DOCUMENT RESUME

ED 087 191 EC 061 293

AUTHOR	Steele, Joe M.
TITLE	Me and My Environment Formative Evaluation Report 2. Assessing Student Abilities and Performance: Year 1.
INSTITUTION	Biological Sciences Curriculum Study, Boulder, Colo.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
PUB DATE	Dec 73
GRANT	OEG-0-9-152075-3720(032)
NOTE	51p.; For related information see EC 050871, EC 050872, EC 050873, EC 050874, EC 050875, EC 061290, EC 061291, and EC 061292
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; Adolescents; Biology; *Curriculum Evaluation; *Educable Mentally Handicapped; *Environmental Influences; *Exceptional Child Education; *Field Studies; Instructional Materials; Mentally Handicapped; Rating Scales; Sciences; Statistical Data; Test Interpretation

ABSTRACT

Presented is the second in a series of formative evaluation reports which summarizes student abilities and performance in field tests of Me and My Environment, a 3-year life science curriculum for 13- to 16-year-old educable mentally handicapped (EMH) adolescents. Discussed are the purpose and interpretation of student data for judging a curriculum. Described in relation to development of test items for the first field test year are aspects such as item format analysis. Functional abilities of students are considered in relation to intelligence and achievement tests, teacher rating of students, problem solving, cognitive development, grouping, and prerequisite knowledge. Explained are differences in performance seen in results of regression analysis and differences in performance among classes. Student performance is analyzed for the directionality and map reading, measurement and scale, temperature, and environmental subtests in Unit I; and for the energy, food chains and webs, food energy, weight and temperature, graphing, and categorizing subtests in unit II. Results are given which show that performance on 25 items in Unit I was not as high as expected, that students in one fourth of the classes showed marked gains on posttests, that one third of students in nine classes were successful on at least one subtest, and that performance on 19 items in Unit II was moderate (one third of students in six classes performed successfully on at least one subtest). Considerable revision of both units is foreseen. Also, findings are given to indicate that individual intelligence tests explain little of the variance in student performance, whereas problem solving and teacher rating do explain the variance. (MC)

Me and My Environment

FORMATIVE EVALUATION REPORT 2

assessing student abilities and performance: year 1

December 1973

Prepared by
Joe M. Steele

BIOLOGICAL SCIENCES CURRICULUM STUDY

ME AND MY ENVIRONMENT FORMATIVE EVALUATION REPORT 2

The research reported herein was performed pursuant to a grant with the Bureau of Education for the Handicapped, U. S. Office of Education, Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position of the Bureau of Education for the Handicapped.

Project No. 1520-75 (Continuation)
Grant No. OEG-0-9-152075-3720(032)

Department of Health, Education, and Welfare
U. S. Office of Education
Bureau of Education for the Handicapped

FORMATIVE EVALUATION REPORT 2

THE CONTEXT FOR THIS REPORT

ME AND MY ENVIRONMENT is a three-year life sciences program developed specifically for 13- to 16-year-old educable mentally handicapped (EMH) children. Its development and assessment, the actual materials and their use in the classroom, the approaches to data collection, and the student outcomes will all be subjects for study. These evaluative activities might best be viewed in the context of the four and one-half year timeline for the development, testing, and final commercial release of ME AND MY ENVIRONMENT. In order to make this curriculum available to special educators as soon as possible, the field trials overlap so that complete field tests of the materials are accomplished in three years.

The following table shows the major stages in the development and evaluation of ME AND MY ENVIRONMENT and working titles of corresponding interim reports which are anticipated:

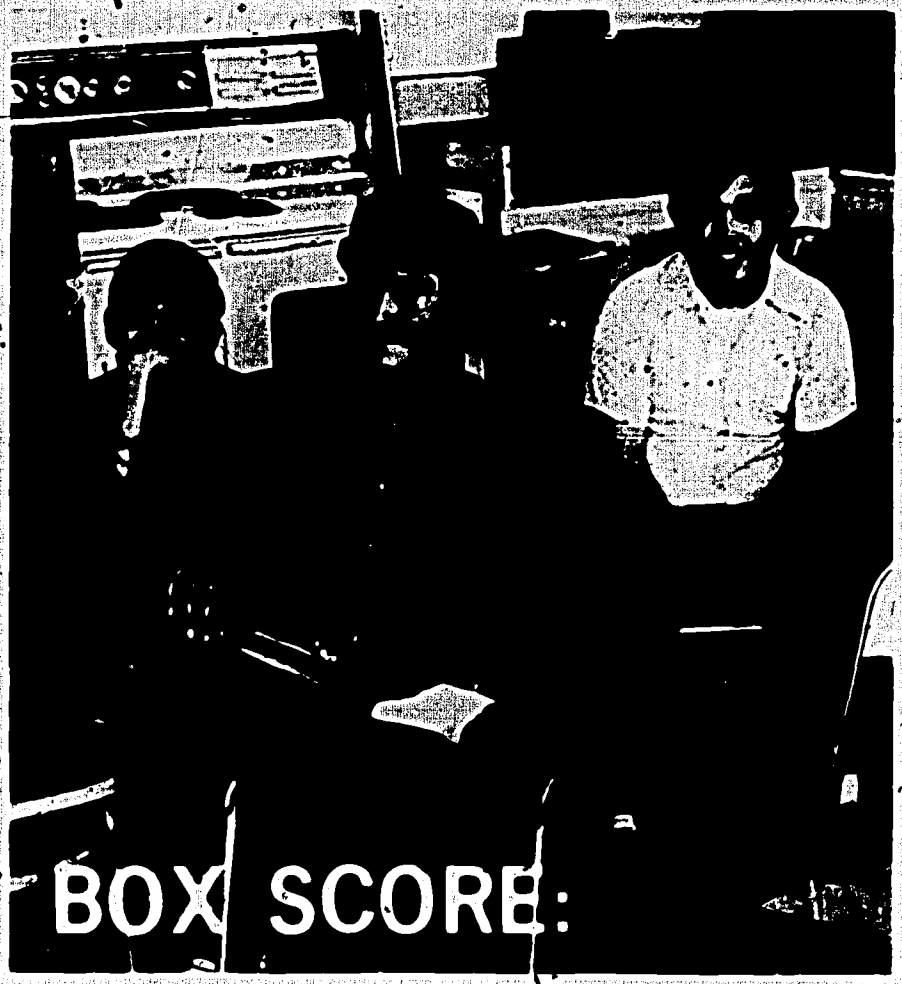
MAJOR STAGES IN THE DEVELOPMENT AND FORMATIVE EVALUATION OF ME AND MY ENVIRONMENT	CURRENT AND ANTICIPATED EVALUATION REPORTS
0. Development of Experimental Materials, Units I-IV (June-October 1971)	0. Plans for Formative Evaluation (Evaluation Issue, BSCS NEWSLETTER 46, February 1972)
1. Content Analysis of Experimental Materials (November 1971-June 1972)	1. Arranging Field Tests: Characteristics of Sites and Students (Interim Evaluation Report 1–June 1973)
2. First Field Test (November 1971-May 1974)	2. Assessing Student Abilities and Performance: Year 1 (Interim Evaluation Report 2–December 1973)
3. First Revision of Units I and II; Refinement of Units III and IV, development of Units V and VI (June-September 1972)	3. Reviews, Revisions, and Data Collection Procedures (Anticipated publication–March 1974)
4. Second Field Test (November 1972-May 1974)	4. The Second Year: Assessing Student Abilities and Performance (Anticipated publication–May 1974)
5. Content Analysis of First Revision Materials (December 1972-June 1973)	5. Assessing Revised Materials: Special Studies (Anticipated publication–August 1974)
6. Second Revision of Units I and II for Commercial Publication (February 1973-January 1974)	6. The Third Year: Assessing Student Success (Anticipated publication–October 1974)
7. First Revision of Units III and IV; Refinement of Units V and VI (June-August 1973)	7. The Formative Evaluation of ME AND MY ENVIRONMENT: A Synthesis of Findings (Anticipated publication of final report–December 1974)
8. Second Revision of Units III and IV for Commercial Publication (February 1974-January 1975)	
9. First Revision of Unit V for Commercial Publication (February 1975-June 1975)	

The materials in the ME AND MY ENVIRONMENT program consist of a series of Teacher's Manuals with suggested teaching strategies for three years of daily science instruction. A kit of all equipment and supplies not ordinarily available in a special education classroom is an integral part of the program and instruction. The materials do not include a student text, as the program is designed around student conducted activities supported by a variety of multi-sensory and media instructional materials. Some of these, in addition to science equipment in the kit, include slides, cassette tapes, individual student worksheets, games, posters, wall charts, illustrated booklets, and evaluation materials. The program makes use of a 35mm slide projector and an overhead projector; active student involvement with a Polaroid Camera and a cassette tape recorder is also being field tested.

The serious reader of this report will likely have reviewed, or have access to, the Teacher's Manuals to ME AND MY ENVIRONMENT. Therefore, information on the objectives, science content, and skill development of the curriculum will not be described here. (Refer to the front material in any unit of the Manuals for this information.) The current project and its evaluation are based upon several years experience in developing and field testing ME NOW, a life science curriculum for 11- to 13-year-old EMH children.[1] The ME NOW program and the first year of ME AND MY ENVIRONMENT are available commercially from Hubbard Scientific Company, Northbrook, Illinois.[2] Several evaluation reports are available on these programs.

[1] *ME NOW, LIFE SCIENCES: A SPECIAL EDUCATION PROGRAM, Biological Sciences Curriculum Study, 1972.*

[2] Hubbard Scientific Company, 2855 Shermer Road, Northbrook, Illinois, 60062.



BOX SCORE:

A SUMMARY OF THE SIGNIFICANT FINDINGS TO DATE

Formative Evaluation Report 1 served four purposes:

1. It defined the function of evaluation in curriculum development and presented the evaluation design for the development of ME AND MY ENVIRONMENT.
2. It described the criteria and procedures for selecting field test participants.
3. It presented data on the actual composition and characteristics of sites and of the students who participated in the first year's field trials.
4. It reviewed the placement of students in this sample of special education classes and drew upon teacher descriptions to portray more fully the variety of young people and instructional problems found in these classes.

In Formative Evaluation Report 2 the following purposes are served:

1. The collection and interpretation of student data is placed within the context of the overall evaluation design.
2. Procedures for development of test items are described and studies of special problems are reported.
3. The development of instruments to assess functional abilities of students is reported, and results for the field test group are presented.
4. Student performance related to instruction in the two units field-tested is reported.
5. A statistical analysis of differences in performance related to a number of variables is presented.

The following represents a summary of the significant findings contained in these two reports.

Identifying the EMH Student

An analysis of the field test population revealed that over one third of the students in the special education classes studied had greatly outdated intelligence test scores. Another third of these students, based on teacher reports, were placed in these classes for reasons other than evidence of retardation. The appropriate placement of 42 percent of the test population could be questioned when one took into consideration tentative evidence of errors in measurement of intellectual functioning coupled with placement of students in these classes for reasons other than retardation. At the same time one notes these problems, it must be pointed out that districts and teachers seemed to have operated with the best of intentions in placing students in these special classes, at least in many cases. Descriptions of students supplied by teachers of these field test classes revealed a wide range of educational and behavioral problems. Placement in these special education classes may well have represented the best available programming for the children involved, given existent funding and legal definitions of these classes. However, the consequences for the children involved are extensive not only in the stigma of being labeled retarded but also in the widening educational gap created by their segregation from other children and from the regular curriculum (see Interim Report No. 1 for supporting data).

Findings in the present report added further concern over the method of identifying EMH children. The individual intelligence test total score appeared to explain little of the variance in performance of students on the life sciences curriculum materials. This finding held for both of the units of instruction taught during the first field-test year. Almost half of the variance in student performance on these materials was explained by a few items assessing problem solving (an aspect of cognitive developmental level) and teacher ratings of students' ability to follow directions or work with their hands. It thus appears warranted to explore other ways of grouping, and even other criteria for placing children in special education classes. The relationship of problem-solving abilities to performance in other subject areas also warrants exploration.

Student Understanding of Science Concepts

This report provides the *first* results of measures of understanding of selected science concepts included in the ME AND MY ENVIRONMENT materials. For this reason the same caution is in order that was made in the first reporting of results on science by the National Assessment Project: "The reporting of the results for the first time in any subject area will not provide a measure of the progress of learning of the population assessed. The first reporting in a given subject provides 'bench-mark data' against which the results of later assessments in that subject can be compared."* The approach to reporting results used herein is similar to that used by National Assessment. Student responses to individual items are shown. Although the efficacy of the curriculum must be judged by results from tests of the revised materials, the general levels of success in ten science areas assessed in the first test are summarized here.

*National Assessment of Education Progress. Report I. 1969-1970 Science: National Results, July 1970, page x.

Subtest 1. Directionality and Map Reading. At least half of the students had a rudimentary grasp of direction, but far fewer could deal with the relationships among landmarks on a map. About half of the students appeared to have a knowledge of compass direction, an understanding felt to be essential in this area of instruction. The standard of 80 percent successful had been set by the staff as a necessary level to judge the instruction acceptable; therefore, considerable revision of the materials has occurred, the level of complexity has been reduced, and elemental knowledge of left-right is being emphasized.

Subtest 2. Measurement and Scale. Rather than assess directly the students' ability to measure things, this subtest was intended to discover whether they had some idea of appropriate units of measure and an approximate concept of feet as a unit of measure. Only about a third of the students could successfully deal with these ideas after instruction. Included in this subtest, also, were items on the concept of scale – that is, approximate representations to scale. Less than a third of the students were successful with these items. On reviewing the materials, the staff eliminated the content on scale as having little relevance for this population of students. Activities on the use of the ruler were revised to begin on a much more elemental level. More practice in measuring and estimating short distances in feet and inches was provided. A minimum standard was set for the second field test: 50 percent of the students should be able to use a ruler to measure short distances accurately.

Subtest 3. Temperature. This subtest dealt with ability to read the temperature on a picture of a thermometer. It also involved an awareness of variations in temperature in the environment, and the concept that dark-colored things absorb more heat than light-colored things. Performance on individual items was surprisingly high, ranging from 65 to 75 percent. Understanding of the *group* of items, however, was considerably lower. Review of the materials resulted in the development of four activities on temperature where one originally existed, and the provision of many new opportunities for students to measure temperatures. The concept of heat absorption has been deleted from the materials.

Subtest 4. Environmental Comparisons. This group of items assessed understandings central to the curriculum materials, although some of the items required going beyond what was included in instruction. Of a total of ten items, most reflected a low level of correct responses on the pretest and marked gains in correct responses on the posttest. From half to three fourths of the students made correct responses to individual items on the posttest. To remedy some weaknesses in the materials revealed by these items, the revised materials contain a great many more opportunities for categorizing, observing, and comparing things. Changes were made in the sequence of instruction relating to the concepts of living-nonliving and life needs.

The preceding four subtests were designed for Unit I of the experimental materials. The remaining six were designed for Unit II.

Subtest 5. Energy. Between 85 and 90 percent of the students had some understanding of the concept that energy is required to do work. From half to two thirds understood that food is the body's source of energy. About 40 percent of the students could predict at which given temperature a liquid could do more work. Two other concepts which were not a part of direct instruction were assessed – a realization that things contain stored energy and the recognition that living things ultimately get their energy from the sun. Less than a fifth of the students understood these ideas. In review, the last two concepts were judged peripheral to the major thrust of the unit and were not developed in the materials. Revisions for the other concepts involved breaking activities into smaller steps, providing more experience with forms of energy, and inserting practical applications of energy concepts. A game was developed dealing with balanced diets and daily food and energy requirements.

Subtest 6. Food Chains and Webs. Half of the students recognized the term *food web* as the correct label for the interrelationships among a pictured group of living things, but less than a fifth of the students could correctly connect a series of pictures to indicate a food chain. For this reason, many changes and additions to the materials were made, resulting in activities dealing with what various animals eat, a revision of a food chain game, a picture booklet dealing with interrelationships among plants and animals, and a series of eight review activities in a later unit.

Subtest 7. Food Energy. About half of the students had some understanding of the idea that all living things ultimately depend on plants as their source of food. Less than half of the students grasped the concept that plants make their own food from sunlight and nonliving materials. To develop these concepts more fully, the sequence of activities in the materials was changed and some of the plant experiments were considerably revised.

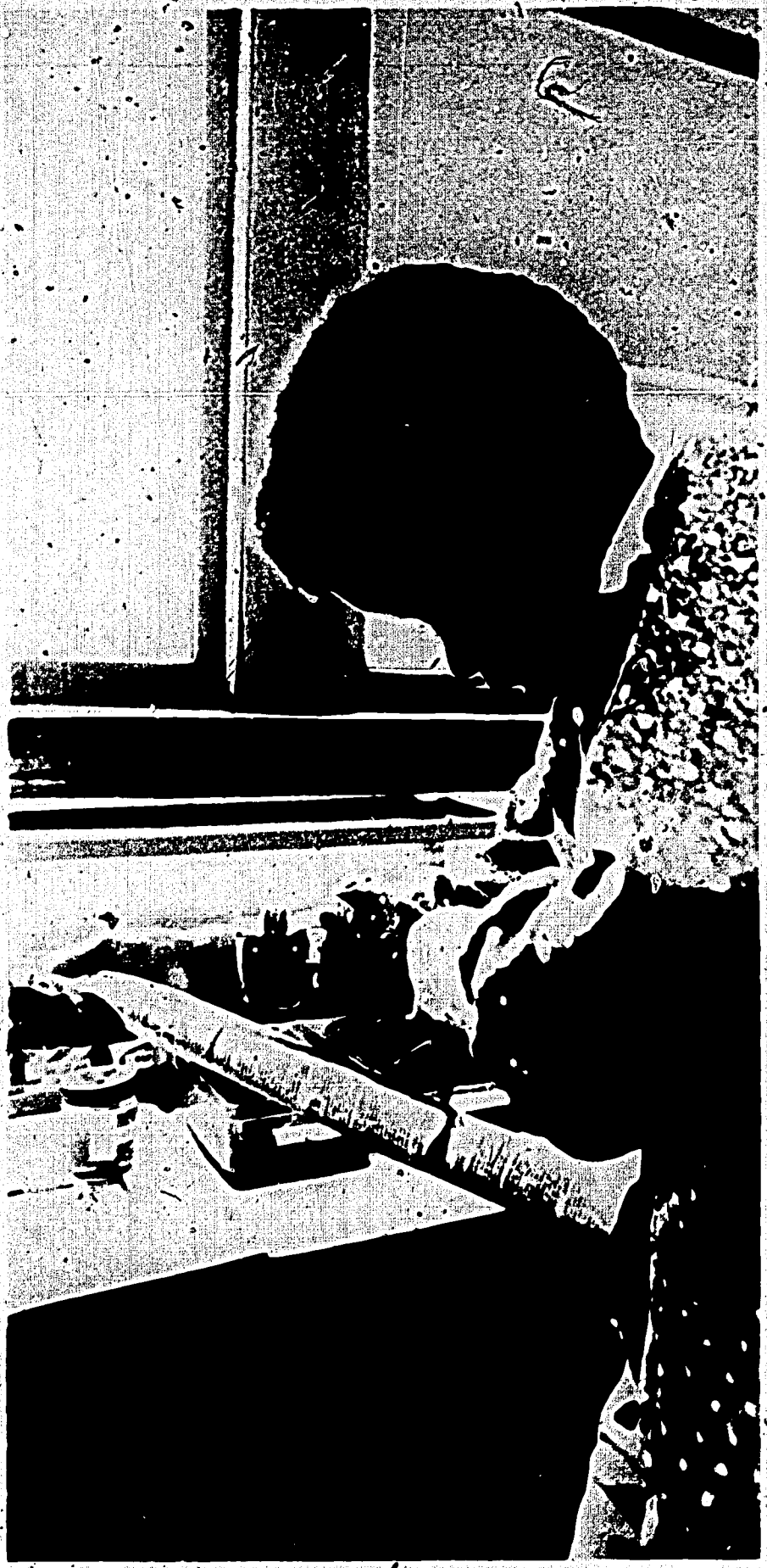
Subtest 8. Weight and Temperature. As in Unit I, a few items were used in Unit II to assess students' understanding of measurement concepts. About two thirds of the students understood how a balance works, and 43 percent of the students recognized 70° as an appropriate room temperature. However, only one third of the students had an understanding of both concepts. In revising the materials, the staff decided to recommend an inexpensive commercial balance and turn the time spent constructing balances to using them. Hence, more time could be spent reading and operating the balance to measure the weights of different things. Knowledge of an appropriate room temperature was not an item of direct instruction; it was assumed that students would have worked sufficiently with thermometers during Unit II to have an awareness of the significance of different temperature levels. Revisions in Unit I have been made to develop the awareness at that point, rather than in Unit II. In the revised materials the balance is also introduced in Unit I.

Subtest 9. Graphing. While several items had been constructed to assess a student's understanding of graphing, defects in some of the items reduced the usable number to one — assessing the student's ability to recognize the correct graphing of three pieces of information. Sixty percent of the students were able to select the correct graph. Because graphs are utilized in various activities in the materials to summarize and compare data, it was felt necessary to devote time to the concept of graphing. A new activity was developed which provides practice in graphing and in reading graphs of various kinds. In addition, the activities which use graphs were expanded to provide more emphasis on the process of graphing itself.

Subtest 10. Categorizing. This subtest consisted of two items that involved not only the process of grouping things but an understanding of specific subject-matter concepts. One item assessed whether children recognized seeds as living things. The other involved comparison of various menus in order to select a balanced meal. About half of the students were successful on these two items. In the revised materials an activity was created in Unit I which calls for students to grow plants from seeds and to maintain plants in the classroom. To further develop the concept of a balanced meal, the "Full and Healthy" game was created for Unit II.

In summary, performance as measured by 25 items related to Unit I of ME AND MY ENVIRONMENT was not as high as the project staff felt was necessary to judge the materials effective. Encouragingly, however, evidences of learning were found in gains from pre- to posttest results on some of the items. In about one fourth of the classes marked gains were shown, and in nine of the field test classes one third or more of the students were highly successful on at least one of the four subtests. The extensive revision of Unit I should result in an increase in the number of students understanding the concepts considered to be most central.

Performance as measured by the 19 items related to Unit II of ME AND MY ENVIRONMENT was moderate. In six of the classes one third or more of the students were highly successful on at least one of the subtests. Considerable revision of Unit II materials should result in both a higher level of success and a greater number of students experiencing success.



THE PURPOSE AND INTERPRETATION OF STUDENT DATA

JUDGING A CURRICULUM

What is the *value* of a given curriculum? Can its efficacy and worth be captured on the first trial by student responses to a few dozen multiple choice items (which are also used for the first time)? We think not. Curriculum development has progressed beyond the "one-shot rocket" material fired at an entire globe of a target with no in-course corrections in aim and no good idea of where it will land — if it hits the target at all.

By the time the ME AND MY ENVIRONMENT curriculum is launched commercially, it will have had many "course corrections." The final revision will barely resemble the first model, a portion of whose testing is reported in these pages.

The value of ME AND MY ENVIRONMENT will be judged eventually by many things:

- --the feelings of teachers using the materials
- --the response and involvement of children
- --the understandings students express in interviews
- --the skills students are able to demonstrate or learn
- --the judgment of experts in the fields of science and special education who review the content
- --the performance of students on test items.

Formative versus Summative Evaluation

Curriculum evaluation at the Biological Sciences Curriculum Study is viewed as an integral part of the developmental process to produce new curricula. The data that serve this purpose must often be gathered in haste and utilized before all the results can be analyzed. This is because the raison d'etre of a curriculum study is to *make available rapidly* its new curricula to schools. Once a product is in use, it can then be assessed in a variety of ways and by a variety of interest groups.

There are many purposes one could wish a single evaluation to serve. It would be desirable to be able to judge the ultimate worth of a program and document the impact that the program has on students during the time the program is being created, first used, and refined. It is understandable that many audiences — legislators, researchers, reviewers — are impatient for such data. It is also not surprising that the functions of *formative* and *summative evaluation* are often confused — these are new concepts, as is much of the theoretical structure of the evaluation field.

However, the primary audience for *formative* evaluation is the developer, whose purpose must be to produce a viable first product. Before *summative* judgments of worth and assessment of outcomes can be made, the existence of a replicable treatment must be established. It is the curriculum developer who is uniquely responsible for providing the data which show that an identifiable program can be successfully installed and operated in a variety of settings. This investigation becomes formative evaluation when the developer not only provides this evidence of a reliable treatment, but identifies defects and weaknesses and then modifies the program to produce the most viable treatment possible. Such is the charge to which the present evaluation design for the EMH life sciences materials is directed. It is in this context that data on student abilities and performance in the first year of field testing of ME AND MY ENVIRONMENT are being reported.

Assessing Student Learning

Many kinds of evidence are being collected that inform us of student learning and abilities. Sources of information include teacher judgments on each activity, observers' notes from the fulltime observation of science instruction in four classrooms, interviews of a random sample of students, and scores from test items and performance tasks. At this point in the four-year process of development and evaluation, these data are tentative and inconclusive. They are fragmented by source and are being reported as the results of each component are processed, rather than being synthesized into a total picture. Such a synthesis would be premature.

The bulk of this report describes the use of over eighty paper-and-pencil test items. These items include questions exploring the background information and experiences of 13- to 16-year-old students, their range of performance in certain general abilities such as grouping, measuring, etc., and their understanding of information provided by instruction. It is easy to put too much trust in scores that can be statistically analyzed. The reader should be cautioned that the data based on student responses are quite tentative. These test items have been used for the first time. New item formats have been tried. Among the instructional items some areas of the curriculum are not adequately represented. The validity and appropriateness of some of the items are still being explored.

What *practical value* does the reporting of this first round of results have? These data serve a number of purposes in spite of the limitations and cautions noted. Several findings may have far-reaching implications for the whole field of special education. Some of the uses to which the data have been put include:

1. Verification of assumptions regarding the background information and skills which students in this age range possess.
2. Exploration of the relationship of general abilities and characteristics of this population to success with this curriculum.
3. Investigation of the range of difficulty and complexity of tasks to which this population can respond.
4. Analysis of student abilities and performance class by class to determine if significant differences in achievement exist.
5. Assessment of student understanding of selected concepts and sequences of instruction.

For the last use these data provide a "bench mark" against which the results from further revisions can be measured. All of these interpretations directly inform the developer of the curriculum. Judgments are possible as to whether curriculum activities begin at the right level and involve tasks in which students can be successful.

Returning to the rocket analogy, these results have more clearly defined the landing site — the target population. Revisions have been initiated to correct the thrust and direction of the materials. The curriculum is aimed at a moving target. Obtaining a "fix" on its position at this point in time has provided a reference point for future checks on the rate and degree of movement of students. Other "in-flight" course corrections have also occurred. Interim Evaluation Report 1 provided an initial analysis of the field test population. Content analysis by specialists resulted in modifications of the "payload." Teacher feedback and staff review also contributed to a number of specific revisions. These are documented in Interim Evaluation Report 3.

THE DEVELOPMENT OF TEST ITEMS

Items Created for the First Field-Test Year

The two units of ME AND MY ENVIRONMENT field tested for the first time during the 1971-72 school year were then called Unit I and Unit II (Part 1). The latter has now been modified to appear later in the curriculum sequence as Unit III.

As a part of field testing the 47 activities in these two units, a total of 83 paper-and-pencil items were created and administered to students. Just as a new curriculum must be tested, so must assessment items and procedures. This article describes four studies which examine problems in using written items for evaluation. Succeeding articles report results for items retained.

In considering the items to be developed, attention was directed not only to assessing concepts included in instruction, but also to evaluating the abilities and knowledge of the students at this point in their development. The items were written and tried out, and after careful study of the validity and functioning of each item, 18 were judged defective and eliminated from further analysis. Of the remaining 65, two thirds (44 items) assessed various areas of instruction and were grouped into ten subtests. One third (21 items) assessed four dimensions of functional ability related to the maturation, cognitive development, and experience of students.

Precedents for Group Testing of EMH Children

Trying to assess the understanding and knowledge of educable mentally handicapped children presents many problems. First of all, some of the children are non-readers. Those who can read encounter much difficulty with some kinds of reading tasks. For example, some of the children have a great deal of trouble following directions. For another example, some have difficulty making decisions based on more than two variables. Many to most of the children have difficulty expressing themselves verbally, especially if this response must be written. Nevertheless, the need existed to assess individuals and groups of students for their understanding of the instructional materials being prepared.

Earlier, during the development and field testing of ME NOW* (the BSCS life sciences curriculum for 11- to 13-year-old educable mentally handicapped children) much effort was expended in developing paper-and-pencil test items which might meaningfully assess student understanding. At length it was demonstrated that these children could respond appropriately to four choices in multiple choice test items which contained a minimum amount of written material. A pictorial or graphic format was used in most of these items. The required response was to mark a particular picture or portion of a picture or to mark a word or short phrase. In that field test each student was provided with a booklet containing the test items. To ensure that the students were always together on the right item, a 35 mm slide of the item was projected. The teacher then read the entire item to the students and gave them time to respond. Hence, reading problems were minimized, the students were kept in the right place, and the teacher could check to see that students were following the directions and marking the items appropriately. As a result, information was obtained from students in group testing situations that required only about 45 minutes per test booklet to administer.

Available commercially from Hubbard Scientific Company, 2855 Shermer Road, Northbrook, Illinois 60062.

Among the things demonstrated was that students could respond meaningfully to this format and did enjoy the testing situation. They were able to attend to the questions for the 30-to-50 minute time period necessary for them to respond to about 30 items relating to a unit of instruction.

The first year of field testing for ME AND MY ENVIRONMENT built upon this method of assessing student performance. A large number of multiple choice, paper-and-pencil test items were developed and tried out (during the 1971-72 school year). These items were grouped into booklets of questions which were administered before and after each of the two units of instruction tested. Since a different age group of children (with respect to those in the ME NOW trials) was involved, they represented in many ways a new population. Therefore, in the first year of field testing an exploration was made to determine the ability of these students to respond to a variety of item formats and to several cuing procedures for keeping their place during administration of the items.

*For a full report see the following:

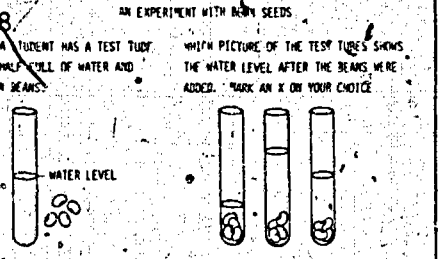
James T. Robinson and Richard R. Tolman. *A Formative Evaluation of ME NOW, Unit I, Digestion and Circulation.* Boulder: BSCS, September 1970. 97 pp. (ED 043182).

Richard R. Tolman. *A Formative Evaluation of ME NOW, Life Sciences for the Educable Mentally Handicapped.* Boulder: BSCS, December 1972, 305 pp. (ED 071263).

Item Format Analysis

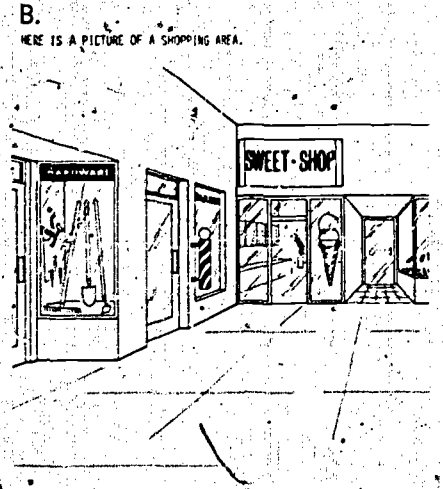
Thirty two questions in booklet form were administered to the first field-test group in November 1971 and again in March 1972. Seven item formats were included. An analysis of responses yielded the following results.

1. One response (marked directly on drawings): Ten items of the type indicated by the following example were used.



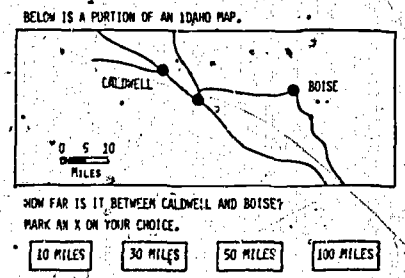
Appropriateness of response was very high. Only one to five percent of the students made no response to one or more of these items. Only one to five percent marked more than one response on any item, except for items 17 and 29, which drew 25 percent and 10 percent multiple responses respectively. These two items were far more abstract and difficult than the other items and received the lowest number of correct responses. Inappropriate responses dropped to two to three percent for most of the items on the posttest. This format was judged to be understandable to EMH children in the 13- to 16-year-old age range.

2. One response (requiring positional mark on drawings): Four items of the type shown in the example were used.



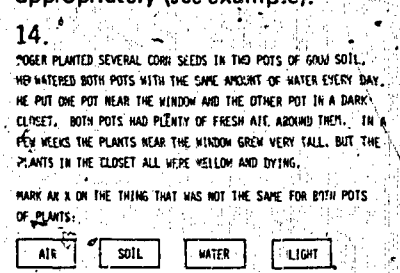
The appropriateness of response was moderately high. Only six to 13 percent of the students made no response to these items. From two to eight percent marked multiple responses. These percentages remained at similar levels on the posttest. The items assessed ability to comprehend and follow directions in order to place the mark in the appropriate position for a correct response. Psychomotor problems related to the ability to position a response accurately did not appear to be a factor in this task. While the format itself did not appear to present problems, these items may have tended to be more difficult than a listing of options from which to choose. The format was judged to be understandable to the target population.

3. One response (one word, phrase, or numeral option): Students showed a high degree of appropriate response to the eleven items using this format. One example follows:



In most cases no more than one to three percent of the students were nonrespondents or marked more than one answer. This format was judged to be understandable to 13- to 16-year-old EMH children.

4. One response (story problems with short responses): The six items of this type have appeared to be among the more discriminating (biserial correlations with total test were higher than for any other items). Yet students encountered little difficulty in responding appropriately (see example):



Multiple responses or no response occurred in two to nine percent of the cases on these items. A similar level of inappropriate responses occurred on the posttest. A pattern of increased multiple response to items which immediately followed others *requiring* a multiple response accounts for most of the inappropriate responses. This format was judged acceptable for use with 13- to 16-year-old EMH children.

5. Multiple responses (one word and picture options): On the four items exemplified by the reprinted item below, the percentage of students who marked a single response was respectively one, four, eight, and 13 percent. On the posttest this level dropped to one percent on three items and five percent on the fourth item. From one to three percent of the students made no response to these items. Thus this format seems appropriate to use with this population.

Before the appearance of the first multiple response item, no more than two to three children marked more than one response per item. On the items immediately following multiple response items, eight to 12 children (seven to ten percent) marked multiple responses — and the same five children accounted for most the perseveration of this multiple response set. Only two other items in the 32-item booklet received multiple markings. While response set may have contributed, it seems likely that the difficulty and unusual format of these two items were the major cause of multiple responses. It therefore was concluded that EMH children in the 13-to 16-year-old age range could handle a combination of formats including multiple response.

(Continued)

Item Format Analysis (Continued)

6. Multiple response (open-ended questions requiring written responses): For each of these three items, three to four words or short phrases were required to be written in, as the example indicates.

NAME _____

WRITE DOWN THE THINGS THAT PLANTS MUST HAVE TO STAY ALIVE.

_____

_____

_____

_____

_____

The level of no response was four to eight percent, which was considered to be quite low. This dropped to one to two percent on the posttest. From four to six percent of the responses were incorrect or undecipherable. Two or more answers to these questions were produced by 57 to 78 percent of the students. Performance on the items was much higher than expected, and increased to 67 to 89 percent after instruction. However, this format does appear to penalize some students and require careful monitoring by the teacher to assure that responses are actually recorded for those students who cannot get their answers into written form. In light of these considerations, a written multiple response is a less appropriate format for this population than the other formats tested.

7. Performance task and written response. Two items required students to sort ten drawings into two groups and name the groups or describe them. (See at right.) These grouping tasks represented a much different format and order of test item. The items were manipulative, applicational, and represented a more direct approach to assessing categorizing ability than a multiple-choice format permitted. In both cases, the children did follow the direction to separate the pictures into two groups. However, from one fourth to one third of the students produced no apparent groupings and were unable to provide labels or descriptions. In a few cases when labels were present, they referred to the top picture in each pile. This result does not imply an inappropriate format any more than it reflects the difficulty of the task. On the posttest, after some experiences in classifying things, the incidence of no apparent groups and no labels dropped to 13 to 14 percent, suggesting that use of performance tasks with brief written responses, while difficult, is not inappropriate for this population of students. Care should be taken to choose tasks which have application to activities students might be called upon to perform.

DATE _____ NAME _____

NAME YOUR GROUPS

ONE GROUP	OTHER GROUP
Task A: Five Pictures of Animals	Task A: Five Pictures of Plants
Task B: Five Pictures of Pollution Scenes	Task B: Five Pictures of Nonpollution Scenes

STAPLE STAPLE

_____ (NAME) AND _____ (NAME)

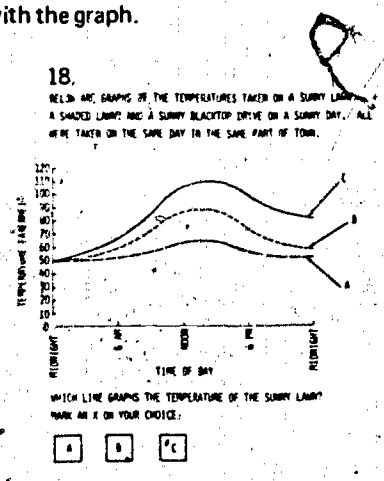
THESE PICTURES ARE ALL IN ONE GROUP BECAUSE _____

THESE PICTURES ARE ALL IN ONE GROUP BECAUSE _____

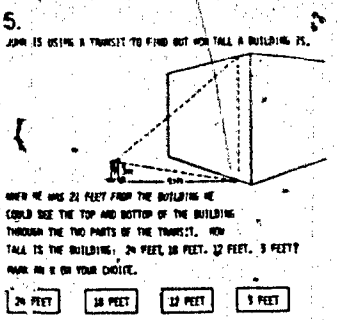
Student Understanding of Items

Things are not always what they seem — and test items are no exception to this truism. Students may interpret items in totally unexpected ways, or not know the meanings of key words. Hence the items may not be an accurate indication of student ability or understanding at all. To the degree that this may occur, the results either become nonsense or must be carefully qualified when interpreted. Thus it is critical to investigate the validity, the meaningfulness, of items.

In the field tests, item validity was explored several ways: by interviewing students, by examining response patterns, and by reviewing the item content in light of instruction and program objectives. Hindsight is sometimes better than foresight, and in review, a few items were judged to involve information that was trivial or peripheral to the main thrust of the curriculum. These items were dropped from the analysis. Several other items proved to contain complex elements which interfered with the assessment of the intended concept. For example, in the following item the students simply could not deal with the graph.



In a few cases, an item appeared straightforward and dealt with concepts that required instruction; puzzlingly, however, most students could answer the item *before* instruction, on the pretest. An example follows.



Some unknown defect in the item resulted in the erroneous indication that most students understood a rather difficult concept prior to instruction.

Some items contained technical defects in art or wording which caused their elimination. The two versions of the following item are examples.

The first version used the term "coolest," which has several connotations. The boys are also doing different things, which led to responses for reasons other than the understanding intended to be assessed. The second version made the boys mirror images of each other, and the question was changed as shown. Of the 47 students interviewed regarding this second version, almost 75 percent marked the correct response. However, 40 percent of those who answered correctly did not have any idea of the concept that white reflects heat, nor any memory of a science activity related to this. Thus, the item still provided an overestimate of the number of students understanding the concept. A more accurate estimate would be that 47 percent of the group sampled grasped the idea.

Student interviews have provided by far the most useful verification of items. About one fourth of the items reported herein were investigated in this way. For example, after the posttesting for each unit of instruction, over half the students in the four Colorado test classes were randomly selected for interviewing. Altogether, over 80 percent of these students were interviewed about some of the items.

The outcomes of item interviews are discussed in later articles, where results of the item are presented. In some cases, substantial validation was obtained, as in the use of the "conservation" item reported under Cognitive Development in the article entitled "Functional Abilities." In other cases, items required qualification of results in light of the degree of guessing or misunderstanding that was evident. In several cases, the interviews revealed factors influencing student learning which would have gone unrecognized. (Examples are reported for Item 25 in Subtest 1 and Item 16 in Subtest 2 in the article "Assessing Student Performance.") As a result, the curriculum has been revised to deal with these issues.

Cues to the Right Question

When ability to read and write is in question, the use of group-administered tests is limited. How can one be sure the student is on the right page and marking an answer to the question being read?

In the ME NOW curriculum, the format used to resolve these problems was 35 mm slides. A slide was made of each item, and each slide was presented on the screen as the teacher read the item to the students. The students matched the item on the screen to the same item in their booklets and thus kept their place in the booklets. This procedure seemed to work quite well.

While the procedure seemed effective, it was also expensive. A study was undertaken to determine whether two less costly procedures for presenting the test items might work as well as the slide procedure, for items to be used with ME AND MY ENVIRONMENT.

(Continued)

a. In the first of these procedures, the items were prepared in test booklets with only one item per page. The booklets were 8½ X 11 inches, and each page was clearly numbered. Each item was read to the class, and after the students had responded they were told, "Turn the page. You should now be at the page that has the number X at the top."
This procedure was based on the assumption that the EMH child, age 13 to 16, can follow this kind of instruction and can read numerals as large as 99.

b. In the second procedure, the items were again prepared in test booklets with only one item per page. The booklets were 8½ X 11 inches. The pages of the booklet were of colored paper, with five different colors arranged in a repeating sequence; e.g., white, pink, green, yellow, blue, white, pink, etc. Each item was read to the class, and after the students had responded they were told, "Turn the page. You should now be at the (color) page."
This procedure was based on the assumption that the EMH child, age 13 to 16 years old, can follow this kind of instruction and can recognize the colors used.

c. A third procedure used 35 mm slides of each item in conjunction with test booklets having one item per page as in a and b preceding. This was essentially the procedure used with the ME NOW tests. The paper was white and each item was numbered. A numbered item was read to the class while shown on the screen. After the class had responded they would be told, "Turn the page. The page that is showing should look like what is on the screen." The next slide would be displayed simultaneously.

Four classrooms were used for the study. The basic design was intended to be repeated measures using posttest only (no pretest administered). The design follows with procedures labeled a (number) b (color), c (slides).

TABLE 1
Intended Design for Cuing Study

	Classroom			
Unit Test	1	2	3	4
I	a	b	c	c
II	c	a	b	b
III	b	c	a	a

The first three unit tests of the curriculum were to have been used for the study. Each classroom would have used one of the formats with each test.

As the field test progressed, it became evident that only two units of instruction could be completed in the year. The design that was implemented was as follows:

TABLE 2
Design for Cuing Study as Implemented

	Classroom			
Unit Test	1	2	3	4
I	a	b	c	c
II	c	a	b	b

A classroom observer monitored each test situation. No class or individual difficulties were noted as more prevalent, or as unique, to a specific procedure. Test instructions for all classes were administered by the same BSCS staff member, to reduce possible teacher influence. Four practice items were used to acquaint the class with the testing procedure at the beginning of each test.

But because two, not three, units of instruction were completed, the students were not exposed to all conditions, and an analysis of results could not utilize analysis of variance. Instead, a correlated "t" test was done for each class between the means of Test 1 and 2 for the class. This required that the analysis be done only for those students who took both tests. It also required that both tests be transformed to a common scale. The latter was accomplished by transforming the scores to a distribution with a mean of 50 and a standard deviation of 10.

Table 3 (below) shows the converted score results and "t" values.

Though there is some dependency on the "t" tests (a high converted mean score in one group would force some lowering of means in the other groups), there is a consistent pattern for the means of the slide procedure to be lower than the means of the other procedures. The difference was significant at the .05 level for Class 1 and nearly so for Class 4.

The study is not conclusive. A balanced research design, as originally intended, could not be used, and the results, to the extent that they can be interpreted, apply only to the use of massed items requiring about an hour to administer. For a smaller number of items, the effects of slide use on focusing attention or arousing interest are not known. While further study is needed, the results do bring into question the necessity for using slides in testing, especially in view of their added cost.

What might account for the indicated nature of the results? One possible explanation might be that seeing an item twice — on a slide and on the printed page — was a partial distractor from full attention to the item in either format. In all cases, however, each item was read twice and ample time was allowed for all to respond. As the opportunity arises, the use of slides for test items will be studied further.

TABLE 3
Standardized Scores and "t" Values for Cuing Study

Unit Test	Classroom 1 (n=15)	Classroom 2 (n=21)	Classroom 3 (n=10)	Classroom 4 (n=13)
I	number $\bar{x}$ = 53.8 s = 8.6	color $\bar{x}$ = 47.1 s = 10.1	slides $\bar{x}$ = 49.7 s = 9.6	slides $\bar{x}$ = 50.0 s = 10.9
II	slides $\bar{x}$ = 49.7 s = 9.1	number $\bar{x}$ = 47.3 s = 8.7	color $\bar{x}$ = 52.7 s = 11.5	color $\bar{x}$ = 55.4 s = 9.1
t	2.47	-.12	-1.04	-2.16

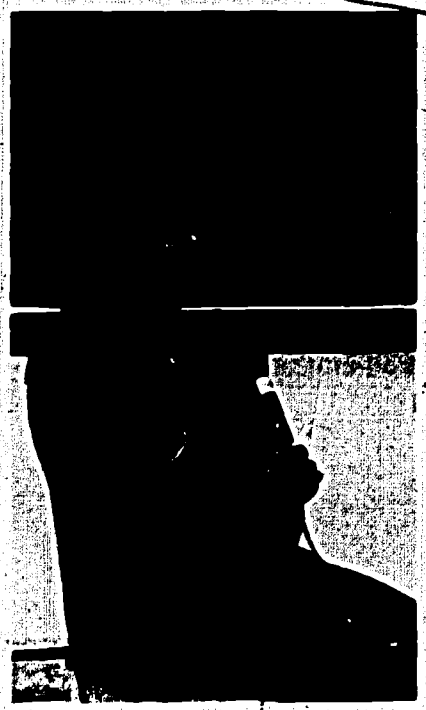
A "t" value greater than ± 2.09 to ± 2.26 is likely to occur by chance only one time in twenty and probably represents al (significant) differences.

Continued Use of Test Items

The use of pre- and posttesting in a formal achievement-testing framework has provided considerable information, as the next several articles indicate. However, it provided a minimum of specific information to writers who were revising the materials, because only a small number of items were tied to individual activities, with most linked to broader sequences of instruction. The items were extremely costly to develop and time-consuming to analyze, and results were incomplete at the time revision took place. In addition, the results were not available in time to be of help to the field-test teachers and thus were of no real value in planning and checking on instruction. Therefore, having obtained data or baseline information on the students and their functional abilities in the first year of field testing, the staff planned a somewhat different approach in order to obtain performance data in the second year of field testing.

Hence, during the spring of 1972, a number of additions and revisions were made to the Teacher's Manual of Unit II to incorporate situational tasks and minitests which would provide immediate evidence of student understanding that would be useful for the teachers in planning instruction or review. These short evaluation sequences were also useful to the project staff in analyzing the performance of activities and making plans for revision. The pilot test of this approach was judged successful, and it was substituted for the pre-post achievement test format in subsequent units and their revisions. Thus, beginning with the second year of field testing, performance data have been collected at the time the concepts were taught. Instructional assessments for short sequences of activities have been incorporated throughout the entire set of materials. These assessments take the form of activities in themselves, and in many cases they involve practical applications or actual performance tasks rated by the teacher. They also include a tallysheet for compiling information on each student or for making ratings of performance. Instructions to the teacher explain how to interpret and use the materials. The tallysheets and in some cases the student worksheets themselves are returned to the BSCS and utilized as the source of data on student performance. For the four units of instruction tested in 1972-73, a total of 131 assessment items was developed, 85 of which can be scored for individual students and the remainder analyzed to make judgments about classroom groups of students. In some cases, items which were used in the first year of testing were incorporated or revised as parts of these instructional assessments.

The assessment activities were given the title "Clues To Success" or "Reviews Of Success." Efforts were made to present them to the teacher not in the light of grading the students but as a means of determining which students were still unsuccessful with the materials and needed additional help. Thus, these assessment activities were designed to establish and help perpetuate a student success syndrome. The data collected will continue to provide the revision team of writers with specific performance data related to each activity or small cluster of activities. If the assessment activities also prove to serve the immediate need of the teacher in planning further instruction, they will remain an integral part of the curriculum materials. The results of this use of test items are now being collected and will be reported in Interim Evaluation Report 5.



FUNCTIONAL ABILITIES OF STUDENTS

Tests of Intelligence versus Functional Ability

Intelligence tests by tradition and purpose are deliberately constructed to measure a relatively stable, global aspect of human potential. They originated out of a need to screen and limit those allowed to continue in school to a well-endowed group who could be predicted to be successful at academic tasks within the period of time typically allotted for instruction in them. Such tests do not provide information that is useful in guiding instruction or in guiding *the educational placement of children.* They simply predict that certain children are likely to be unsuccessful in performing the academic tasks traditionally demanded by schools.

The other major instrument for assessing school children, achievement tests, is also designed to accentuate the differences among children rather than provide a guide for their instruction. It, too, is a screening and sorting device in the tradition aimed at the production of scholars. The children with which such a system cannot deal are routed through watered-down versions of the same academically oriented curricula or are the children who make up the special education classes in the schools.

The problem the teacher faces in any classroom, but critically in the special education class, is to discover what each child's level of functioning at a given time actually is. Only then can the learning materials be mediated to accommodate the level and rate of functioning of each child. Note that functional ability is assumed to change as a child develops and to be specific to the kind of task rather than a global measure. It does not indicate potential, but present capabilities. Unfortunately, actual tests of functional ability do not exist. A few are currently being developed. Little is known of the relationship of various abilities to instructional materials in any discipline. It behooves each curriculum developer to explore the abilities judged uniquely requisite for success on specific materials and to provide the teacher with tools to make the success possible, for each student.

The purpose of the BSCS in studying science for children in special education classes is to pioneer a curriculum for doers rather than scholars. Reading, writing, and arithmetic are not the central features of this program. Instead, students do activities and experience situations to gain a practical understanding of the world around them. The materials are intended to be both functional and intellectually stimulating, but in the special education setting this means calling on a different set of skills. What *are* these skills for the doers, that will enable them to cope with practical problems and situations in and out of school? What abilities influence the acquisition of competence in various tasks? And at what levels of functioning are the children to be found in their special classes? These are concerns of the project staff in the development of ME AND MY ENVIRONMENT. An aspect of the evaluation has been the attempt to identify some of the functional abilities and to explore whether they are indeed influential in the success of children using the ME AND MY ENVIRONMENT materials.

Even as the staff set about trying to develop some measures of functional abilities, the schools involved in the field tests were requested to furnish current individual intelligence test data. The point was to determine whether this instrument so widely used in our culture and required by law for placement in special classes provides educationally relevant data for this population. The next article deals with such an analysis.

Teachers were asked to rate students on a number of criteria which were judged to contribute to successful performance of the science activities. These included:

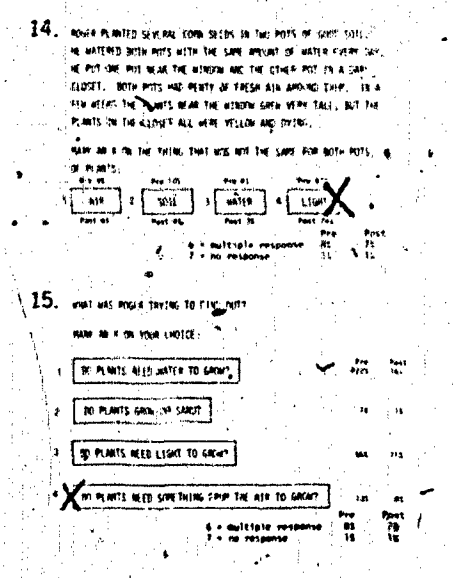
- --verbal participation by the students
- --their ability to follow directions
- --their ability to work in a group
- --their ability to attend to a task for a sustained period
- --their ability to work with their hands
- --their general attitude toward school
- --awareness of things going on around them
- --the reason for placement in a special class.

More is presented about these ratings and how they affect success in the next article. See the inset for a description of the rating form and each set of categories as responded to by the teachers.

To assess some functional abilities more directly, a number of tasks were developed to which students responded. The tasks ranged from directions to measure something or group things to marking multiple choice items. The tasks or items were divided into four subtests of functional ability: Problem Solving, Cognitive Development, Grouping, and Prerequisite Knowledge. While the items had some relationship to the science materials, it was assumed that they represented experiential and developmental dimensions which were unlikely to be affected by instruction of relatively short duration. The following sections contain descriptions of these items and results of use. Following these descriptions and data, the next article presents information on relationships of these abilities to performance.

Problem Solving

Six items assessed critical aspects of problem solving in the context of conducting an experiment; all are reproduced in this section. Two of the items dealt with knowing the experimental conditions (Items 14 and 20), two dealt with knowing the question to be answered in an experiment (Items 15 and 30), and two dealt with recognizing conditions which might influence the results of an experiment (Items 29 and 31). That the items are clearly related in assessing a common skill labeled "problem solving" is indicated by the fact that the R-biserial correlations of individual items with the total score on this subtest were above .90 for four of the items, .81 for a fifth item, and .63 for the sixth item (NOTE: maximum values for an R-biserial are + 1.25). Levels of correct response to the items were above two thirds of the total group for all but two items. On one of these two items, 58 percent of the total group correctly answered one of the items dealing with knowing the experimental conditions. The lowest percentage of correct response (37 percent) was made on one of the items dealing with recognizing conditions which might influence the results of an experiment. Since the students did experiments, these items certainly were related to instruction. In several cases, classroom groups showed large gains on most of the items, as the chart of pre- and posttest scores, by class, indicates. Nevertheless, it was felt that these items tapped abilities more related to developmental and experimental dimensions than to achievement. As the multiple regression analysis in the next article shows, the problem-solving subtest explained almost 40 percent of the variance in performance on the 44 items used to assess instruction.



Considering the same six items as a subtest, almost one fourth of the students could be judged low in problem-solving ability, answering two or fewer items correctly. This is one standard deviation below the mean for the group. One fifth of the students were high in this ability, answering all six of the items correctly.

Because of the relationship between problem solving and success on the instructional items, special attention has been given to the development of problem-solving skills in the revision of the materials.

TABLE 4
Problem Solving Subtest

Subtest Weighted Scores:

Item	Weights Per Option 1	2	3	4	5	6	7
14	0	0	0	5	-	0	0
15	0	0	5	0	-	0	0
20	0	0	5	0	-	0	0
29	0	0	5	0	0	0	0
30	5	0	2	0	-	0	0
31	0	5	2	0	-	0	0

Subtest Statistics:

Maximum possible score: 30
Range of Scores: 0-30
Mean: 19.3
Standard Deviation: 9

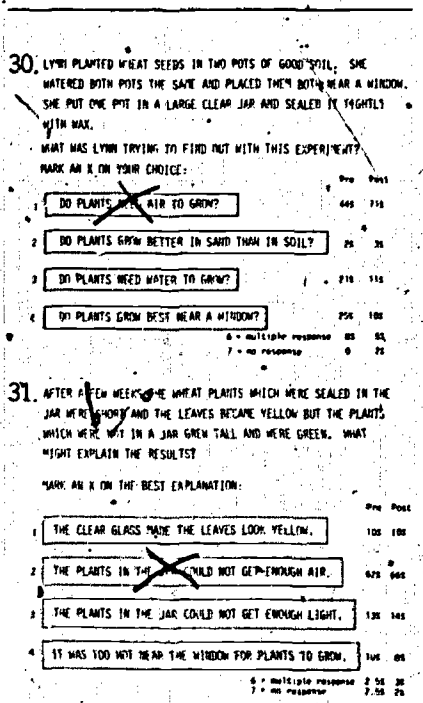


TABLE 5
Problem-Solving Ability: Percentage Choosing Correct Response for Each Item by Class and Total Group*

Questions N	00 Pre	00 Post 14	05 Pre	05 Post 4	06 Pre	06 Post 9	07 Pre	07 Post 10	08 Pre	08 Post 9	09 Pre	09 Post 9	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Pre	All Classes Post 109
14	54	75	90	64	78	89	77	76	60	65	58	83	79	76	93	68	67	78
15	46	81	50	64	44	89	77	65	50	67	58	50	64	71	86	44	56	71
20	38	60	30	67	20	80	29	80	50	50	58	73	55	40	85	48	38	58
29	31	27	70		20	40	25	73	58	60	58	36	27	40		17	48	37
30	54	81	20	73	44	78	54	76	30	67	50	67	64	71	79	44	44	71
31	46	75	73	55	67	67	62	76	40	42	83	50	79	71	71	48	62	66

Posttest only Subtest Performance:**

Low (two or fewer items correct)					33		44			52		24
High (all items correct)			44	40								20

*Complete data were not obtained from classes 01, 02, 03, and 04. For purposes of the cuing study described in the previous article, pretests were not administered to classes 11, 12, 13, and 14.

**Percentages of students are shown for each class in which more than one fourth of the group scored more than one S.D. from the mean.

Cognitive Development

Three items were developed to assess the ability to deal with logical and abstract relationships in contrast to concrete, tangible manipulations of objects. All three are reproduced in this section.

These items are related to Piaget's theory of cognitive development. One (Item 1), on conservation of liquid quantity, involved the recognition that liquid quantity does not change with the size and shape of the container. A second (Item 17) involved the ability to serially order objects from smallest to largest. The third (Item 33) dealt with the ability to translate a three-week interval of time into a specific calendar date given the beginning date and the calendar on which to make the calculation. Evidence that these three items are appropriately grouped to measure the concept of cognitive development is the high *R*-biserial correlations between each item and the total group score on this subtest. (The R-biserial correlation for the conservation item is 1.00, for the calendar item .98, and for the seriation task .54.) Normally, the use of tasks like those used by Piaget requires individual interviews, where the student explains his ideas after he has performed the task. Because this approach is prohibitively expensive, the BSCS conducted a study to determine whether a teacher-demonstrated paper-and-pencil test for assessing conservation of liquid quantity might prove equally valid. The sample for this study utilized students from the four field-test classes in Colorado. The day after the written test was administered, 47 of the 70 children were randomly selected and individually interviewed as each performed the Piaget task in the traditional manner. Orange juice was used, and, after the child performed and explained the task, he was asked whether the beaker or the test tube contained more orange juice for him to drink. The student's explanation of his choice supported his other responses.

Results were as follows: One of the 47 children was inconsistent in the interview; on his test item he marked "They both have the same amount of water," but he could not conserve liquid quantity for orange juice in the interview. Three children showed transitional responses; they had marked an incorrect option on the test item but appeared to grasp the idea during the interview. The responses of the remaining 43 children, 92 percent of those sampled, were consistent with the way they had marked the test item. These results suggest that, *for this population,* it is possible to assess at least some aspects of cognitive development by using written items in a group setting.

For the total group of students in

(Continued)

Cognitive Development (Continued)

the field test, 54 percent marked a response indicating the ability to conserve liquid quantity; 55 percent were successful with the seriation task. However, only 18 percent of the students could successfully translate the interval of time into a calendar date. Considering the three items as a subtest, 15 percent of the students were high performers, answering more than two items correctly, while 17 percent were low performers, answering none of the items correctly. The high and low performers were distributed widely across all of the test classes in the sample.

TABLE 6
Cognitive Development Subtest

Subtest Weighted Scores:

Item	Weights Per Option 1	2	3	4	5	6	7
1	5	0	0	0	-	0	0
17	1	1	1	5	3	0	0
33	0	0	3	5	0	0	0

Subtest Statistics:

Maximum possible score: 15
Range of Scores: 0-15
Mean: 6.3
Standard Deviation: 4.3

TABLE 7
Cognitive Development: Percentage Choosing Correct Response for Each Item by Class and Total Group*

	00		05		06		07		08		09		11	12	13	14	All Classes	
Questions / N	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Post	Post	Post	Post	Pre	Post
	14		4		9		10		9		9		9	15	11	19	109	
1	77	50	50	27	33	67	62	76	30	42	92	67	50	76	21	48	57	54
17	13	27	20	50	70	70	65	53	42	50	67	82	73	47	46	52	54	55
33	13	13	10	17	0	50	12	20	33	30	8	0	18	0	23	9	20	18

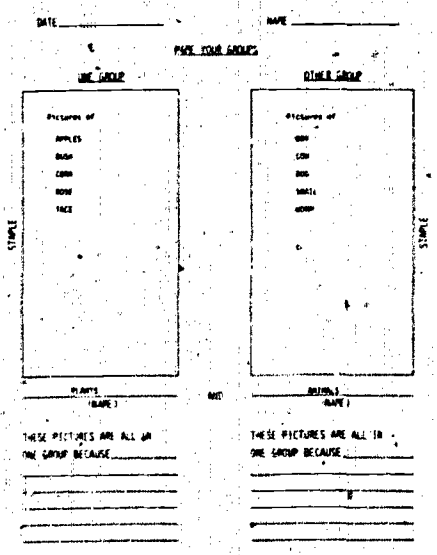
Posttest only Subtest Performance:**

Low (none correct)	35				44						26	17
High (more than two items correct)		25	33	30	44		33					15

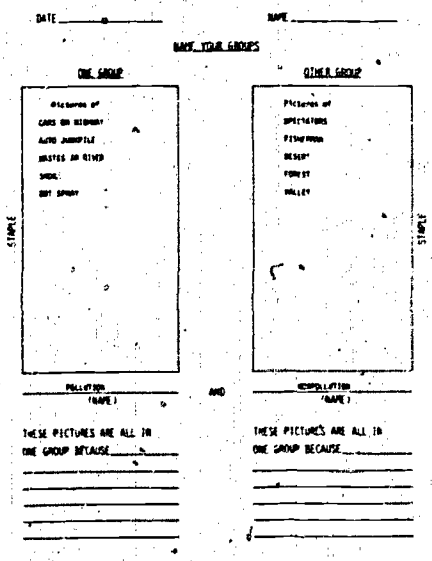
*Complete data were not obtained from classes 01, 02, 03, and 04. For purposes of the cuing study described in the previous article, pretests were not administered to classes 11, 12, 13, and 14.

**Percentages of students are shown for each class in which more than one fourth of the group scored more than one S.D. from the mean.

35



36



Grouping

Six items were designed to assess the ability to sort and classify objects according to their common characteristics. They are included in this section. Two of the items were sorting tasks (Items 35 and 36). The other four were paper-and-pencil items (Items 17, 37, 38, and 39). All but one of the items, however, were performance tasks rather than strictly selection or recognition. Most required the ability to place objects in appropriate groups, as well as the assignment of reasonable labels to such groups. All reasonable groups were accepted in the scoring procedures. From 44 to 61 percent of the students were able to answer correctly each of the items. One of the items (Item 17) involved recognition of a group of objects sorted on the basis of a specific characteristic. Three other items (Items 37, 38, 39) involved sorting a pictured collection of objects into different groups by changing the characteristic used to group them. The remaining two items (Items 35 and 36) involved sorting a set of ten pictures into two groups and labeling, or describing, the groups. This latter pair of tasks was scored both on appropriately assorted groups and on assignment of reasonable labels. The results indicated that if students were able to identify reasonable groups, they were also generally able to supply a label or description of those groups. Considering the six items together as a subtest, 16 percent of the students were high performers, answering more than five of the items correctly; 16 percent were low performers, answering fewer than three of the items correctly.

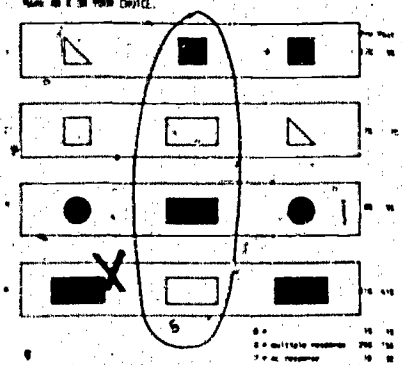
TABLE 8
Grouping Subtest

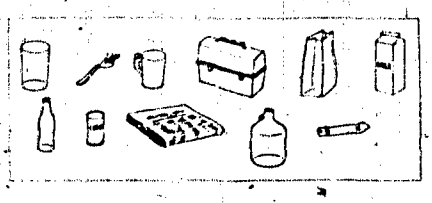
Subtest Weighted Scores:

Item	Weights Per Option 1	2	3	4	5	6	7
17	0	0	0	5	5	0	0
35	0	2	3	4	5	5	-
36	0	2	3	4	5	5	0
37	0	0	3	5	5	-	0
38	0	0	3	5	5	0	0
39	0	0	3	5	5	0	0

Subtest Statistics:

Maximum possible score: 30
Range of Scores: 0-30
Mean: 17.3
Standard Deviation: 8.3





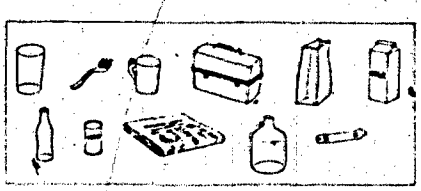


TABLE 9

Grouping Ability* Percentage Choosing Correct Response for Each Item by Class and Total Group*

Questions N	00 Pre	00 Post 14	05 Pre	05 Post 4	06 Pre	06 Post 9	07 Pre	07 Post 10	08 Pre	08 Post 9	09 Pre	09 Post 9	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Pre	All Classes Post 109
17	77	69	0	55	56	11	54	53	50	25	33	58	64	53	50	52	41	61
35**		50		63		22		53		33		67	64	41	42	48	35	53
36**		19		36		44		59		58		75	65	58	57	40	24	52
37	25	60	20	33	40	50	29	53	33	40	25	45	46	20	69	48	39	47
38	56	47	20	58	30	60	29	73	58	60	33	36	27	20	54	39	42	45
39	56	60	30	58	40	60	29	66	42	50	25	18	36	40	53	35	42	48

Posttest only Subtest Performance***

Low (fewer than three items correct)						33			31		16
High (more than five items correct)	50		30								16

*Complete data were not obtained from classes 01, 02, 03, and 04. For purposes of the cuing study described in the previous article, pretests were not administered to classes 11, 12, 13, and 14.

**Items 35 and 36 were administered in posttest only.

***Percentages of students are shown for each class in which more than one fourth of the group scored more than one S.D. from the mean.

Prerequisite Knowledge

Six items assessed background information which was not a part of the instructional sequence, but which indicated the kinds of experiences or understandings that might contribute to success with the materials. All these items appear with this section.

The areas of prerequisite knowledge sampled by this subtest included two items (Items 4 and 16) assessing awareness of basic units of linear measure. Thirty one percent of the students answered correctly the number of feet in a yard; 83 percent answered correctly the number of inches in a foot. Another item (Item 38) assessed the ability to use a ruler to measure lengths under 12 inches within ¼ inch accuracy. The level of performance on this task was uniformly low, with only 35 percent of the total group of students succeeding. Still another item (Item 43) assessed the use of feet as an appropriate, standard unit of measure in estimating distance; 42 percent of the students were successful. Finally, two items (Items 13 and 22) assessed awareness of an appropriate range of human body temperature. Forty six to 49 percent of the students had this knowledge. Considering the six items as a subtest of background information, 17 percent of the students were able to answer more than four of the items correctly, while 20 percent answered fewer than two of the items correctly. These items were used as an indicator of the breadth of student background information related to several forms of measurement, rather than for the specific facts contained in the items themselves. Hence, the more experience students may have had in dealing with length and temperature, the more likely (or so it was assumed) that they would be able to answer these questions. The students in this field test were generally more familiar with inches than feet, but few could accurately use a ruler. In half of the classes students had a very poor concept of the size of various standard units of measure and of which unit was appropriate in a given situation.

This exploration revealed that both the curriculum developers and the teachers in the field test were making unwarranted assumptions about the students' ability to accurately use measuring instruments. Therefore, the revised activities were designed to start at much more basic levels in dealing with the use of measuring devices, and in making measurements and estimates.

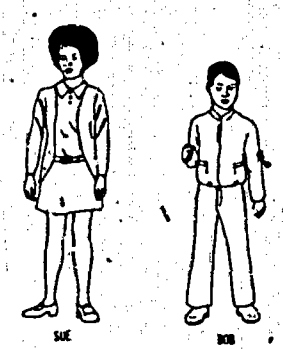
TABLE 10
Prerequisite Knowledge Subtest

Subtest Weighted Scores:

Item	Weights Per Option 1	2	3	4	5	6	7
4	0	5	0	0	-	0	0
13	0	0	5	0	-	0	0
16	0	0	5	0	-	0	0
22	0	0	5	0	-	0	0
38	0	0	0	3	5	0	0
43	5	0	0	0	0	0	0

Subtest Statistics:

Maximum possible score: 30
Range of Scores: 0-30
Mean: 13.9
Standard Deviation: 7.4



13.
SUE AND BOB ARE TWO HEALTHY CHILDREN. SUE'S TEMPERATURE IS 98.6 DEGREES. WHAT DO YOU THINK BOB'S TEMPERATURE IS?
BOB'S TEMPERATURE IS _____

Pre	Post	
13%	75%	1 = much higher or lower temperature
24%	14%	2 = a little above or below (98 + 102)
40%	45%	3 = 98.6°, 98, the same
5%	14%	4 = wrong or inappropriate response
0	2%	8 = multiple response
9%	4%	9 = no response

22.
IF YOU HAVE A FEVER, THE THERMOMETER WOULD SHOW ABOUT WHAT TEMPERATURE?

MARK AN X ON YOUR CHOICE.

	Pre 17%	Pre 17%	Pre 46%	Pre 20%
	72° F.	98° F.	102°	112° F.
	Post 13%	Post 18%	Post 49%	Post 20%

	Pre	Post
8 = multiple response	0	2%
9 = no response	0	1%

4.
HOW MANY FEET ARE IN A YARD?

MARK AN X ON YOUR CHOICE.

	Pre 3%	Pre 27%	Pre 23%	Pre 47%
	2	3	12	36
	Post 7%	Post 31%	Post 19%	Post 50%

	Pre	Post
8 = multiple response	0	0
9 = no response	0	0

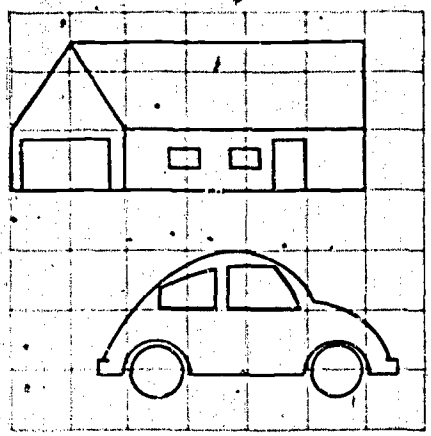
16.
HOW MANY INCHES ARE IN A FOOT?

MARK AN X ON YOUR CHOICE.

	Pre 5%	Pre 9%	Pre 74%	Pre 13%
	2	3	12	36
	Post 1%	Post 4%	Post 84%	Post 10%

	Pre	Post
8 = multiple response	0	1%
9 = no response	0	1%

38. (READING) PICTURE



1. WHAT IS THE LENGTH OF THE HOUSE ON THE PICTURE? 6 inches (within 1/8")
2. WHAT IS THE LENGTH OF THE HOUSE ON YOUR DRAWING? 3 inches (or within 1/8")
3. WHAT IS THE SCALE OF YOUR DRAWING COMPARED TO THE PICTURE? _____
4. HOW TALL IS THE HOUSE ON THE PICTURE? 2 1/2 inches (within 1/8")

Post only

Code

1 = Only question 1 correct (within 1/8")
2 = Only question 2 correct (within 1/8" whether drawn to scale or not)
3 = Only question 3 correct (within 1/8")
4 = Both 1 and 2 correct
5 = Questions 1, 2 and 3 correct
6 = Incorrect responses to all three questions
7 = No response

43. GUESSING GAME (ESTIMATING DISTANCES)

TABLE 11

Prerequisite Knowledge: Percentage Choosing Correct Response for Each Item by Class and Total Group*

Questions N	00 Pre	00 Post 14	05 Pre	05 Post 4	06 Pre	06 Post 9	07 Pre	07 Post 10	08 Pre	08 Post 9	09 Pre	09 Post 9	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Pre	All Classes Post 109
4	13	40	30	17	20	30	35	53	25	30	42	36	36	33	31	9	27	31
13	44	47	50	58	60	90	18	40	58	70	50	64	36	53	46	22	48	46
16	69	80	70	67	70	80	65	100	67	80	67	73	91	87	85	78	74	83
22	50	60	30	42	40	80	41	47	42	30	42	64	55	40	62	35	46	49
38**		6		0		22		12		8		33	7	12	7	16		35
43**		0		73		67		47		100		92	29	29	29	20		42
/ Posttest only Subtest Performance***																		
Low (fewer than two items correct)		28		50										26		47		20
High (more than four items correct)						33						44			27			17

*Complete data were not obtained from classes 01, 02, 03, and 04. For purposes of the cuing study described in the previous article, pretests were not administered to classes 11, 12, 13, and 14.

**Items 38 and 43 were administered in posttest only.

***Percentages of students are shown for each class in which more than one-fourth of the group scored more than one S.D. from the mean.

EXPLAINING DIFFERENCES IN PERFORMANCE

What variables account for differences in student performance? According to the analysis reported in this article, none of the traditionally used variables such as age, IQ, or ethnic group provides insight into why students were successful or unsuccessful. Half of the variation in performance is accounted for by three functional abilities (problem solving, grouping, and prerequisite knowledge) and two teacher ratings (following directions, and ability to work with hands). Such results are dramatic because they suggest the possibility of predicting which students will be most or least successful with the materials. These variables also focus attention on skills and activities which may lead to improved student performance.

How believable are the results? Those readers who would like a technical report should refer to the following section describing results of the statistical analysis. Essentially, it reports that problem solving, a six-item test of functional ability, explained twice as much of the variance in student performance as did IQ scores, and that this result was repeated for performance on both units of instruction. The questioning reader should also examine the kinds of test items from which a performance score was derived. The next two articles present and discuss the items used to assess understanding of the two units of ME AND MY ENVIRONMENT that were field-tested.

What are the implications of these findings? The results seem to bear on both the appropriateness of the curriculum being tested and the manner in which this population is identified and grouped. Because many children achieved success regardless of a wide range of IQ, age, varying ethnic background, and sex, it would appear that the materials are on target. It is especially noteworthy that IQ, which typically correlates highly with success in the traditional academic curriculum, is not a predictor in this intentionally nonacademic, activity-oriented curriculum. Since the curriculum also attempts to develop further the abilities which do explain differences in success, the materials would appear to be especially appropriate. Another implication of these results is to raise the question about the appropriateness of an elaborate and costly system of individual intelligence testing, in light of the finding that intelligence scores are not predictive of performance with this curriculum. Considering the inefficiency with which such a screening system is run, and the large margin for error (as reported in Interim Evaluation Report 1), there is all the more reason to reexamine the entire procedure for placement of children with educational problems. Finally, as intelligence testing does not supply diagnostic information to guide instructional decisions, there is more than sufficient justification to attend to measures of functional ability.

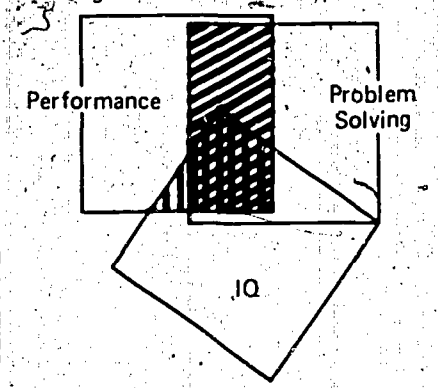
Were some teachers more effective than others in teaching these materials? The last section of this article describes the results of an analysis of differences in performance by class, rather than among individual students, reflecting the teaching ability in the separate classes. After statistically adjusting student scores to compensate for differences in ability, the staff found two teachers (06, 12) to be far more effective than others in teaching one or both units of instruction, and three teachers (08, 09, 13) to be far less effective than others. A major characteristic which might have accounted for these differences was the fidelity with which the Teacher's Manual was used in implementing the materials as intended. Other than this, the differences probably do reflect overall differences in teaching ability.

Results of Regression Analysis (Continued)

If problem solving assessed the same thing as IQ, then forcing IQ to enter the regression equation first would result in most variance being explained by IQ, with little or none of the remaining variance explained by problem solving. Table 13 reports the results of this second analysis: the two tests are not measuring the same thing. While IQ appeared to account for about 20 percent of the variance in performance for both units in the second analysis, problem solving explained an additional 21 percent or more of the variance even with IQ entered first. The interrelations of IQ, problem solving, and performance might be illustrated graphically as follows:

FIGURE 1

Schematic Illustration of Variance Accounted for by IQ and Problem Solving



Shaded areas indicate variance in performance accounted for by the two measures. Problem solving accounts for most of the variance explained by IQ. Hence, the results of the first regression analysis are appropriate to use.

Table 14 shows the interrelationships of the significant variables. The four functional abilities and four teacher ratings are only moderately related to each other. The variance in student performance explained independently by each of these variables is also shown in the first two columns of this table.

TABLE 13
Second Analysis: Variance Accounted for in Regression Analysis, with IQ Forced to Enter First

(F values significant at .05)

Variables	Step*	Unit I	Step*	Unit II
IQ (Total Score)	1	16.8%	1	21.7%
Problem Solving	2	22.2%	2	20.9%
Following Directions	3	8.4%	----	----
Ability to Use Hands	----	----	3	6.3%
Absence Ratio (Unit I)	4	4.3%	9	.9%
Unit I Performance	----	----	4	2.4%
Prerequisite Knowledge	5	3.6%	----	----
Grouping Ability	----	----	5	1.6%
Unit II Performance	6	1.6%	----	----
Test Class	7	.6%	7	1.8%
Reason for Placement	8	1.0%	6	1.3%
Chronological Age	9	.7%	8	1.1%
Conservation of Quantity	10	.7%	----	----
		59.9%		58.0%

*Order in which the variables entered the regression equation.

TABLE 14
Intercorrelations of Significant Variables
(The Variance Explained Independently by Each Variable is Shown in Parenthesis)

			Functional Ability				Teacher Ratings					
Variable	1(r^2)	2(r^2)	3	4	5	6	7	8	9	10	11	12	13
1. Unit I Performance	1.00	.57	.41	.61	.50	.44	.42	.52	.25	.25	.29	-	-.23
2. Unit II Performance	(33%)	1.00	.47	.62	.43	.49	.46	.42	.41	-	.29	.21	-
3. IQ	(17%)	(22%)	1.00	.47	.40	.41	.34	.24	.21	.31	-	-.33	-
4. Problem Solving	(37%)	(39%)		1.00	.48	.45	.49	.40	.24	-	-	-.18	-
5. Prerequisite Knowledge	(25%	(19%)			1.00	.36	.47	.27	-	.35	.34	-	-
6. Grouping Ability	(19%)	(24%)				1.00	.38	.33	.27	.19	-	-	-
7. Cognitive Development	(18%)	(21%)					1.00	.34	.33	-	.27	-	-
8. Following Directions	(27%)	(17%)						1.00	.40	-	.20	-	-
9. Ability to Use Hands	(12%)	(17%)							1.00	.24	.46	-	-.22
10. Reason for Placement	(6%)	(1%)								1.00	.22	-	-
11. Awareness	(8%)	(8%)									1.00	-	-
12. Chronological Age	(.1%)	(4%)										1.00	-
13. Absence Ratio	(5%)	(.2%)											1.00

Results of Regression Analysis

As an attempt to discover factors which might have influenced the performance of students on these materials, a stepwise multiple regression was undertaken. This analysis was conducted both on the 25 instructional items related to Unit I of ME AND MY ENVIRONMENT and on the 19 instructional items related to Unit II. Total scores were derived for each student based on the weights indicated in the following articles describing student performance data. Omitted from the analysis were all students for whom complete data was not available. Some of these students did not complete one of the two Unit achievement tests, or no background data was available for them. The number included in the analysis was further reduced by those students who dropped out or who entered the class during the year, and for whom only partial data was available. Ten classes had completed the second unit of instruction; the number of students included in this analysis was 109. Based on the end-of-year enrollment of 142 children in these classes, 70 percent of them were included in the analysis.

Twenty variables were entered into the regression analysis. These included the score on the other posttest, the experimental class, and the four subtests assessing functional abilities of students (problem solving, cognitive development, grouping, and prerequisite knowledge). One item from the cognitive development subtest, which assessed ability to conserve liquid quantity, was also entered separately. Other variables included sex of the students, chronological age, ethnic group, and Wechsler Intelligence Scale for Children (WISC) total IQ score. The ratio of class absence to the total number of days of instruction was included for each student for Unit I.

Finally, eight teacher ratings of students were entered as variables. Each rating scale had been carefully defined to teachers and included: verbal participation of the student in class; ability to follow directions; ability to work in a group; ability to attend to a task; ability to work with one's hands; general attitude toward school; the student's awareness of subtleties in what was going on around him; and the primary reason for the student being placed in a special education class. (See the section in the preceding article defining these teacher rating scales and the measures of functional ability.)

Table 12 summarizes the variables accounting for most of the variance in the first analysis against Unit I and Unit II performance. In these regression equations the strongest variable was allowed to enter first. This proved to be problem solving, which accounted for almost 40 percent of the variance in performance in both units of instruction.

In order to avoid misinterpreting the results, a second regression analysis was conducted (since the nature of this statistic is that when two measures of the same thing are included, all of the variance will be attributed to the variable entering the equation first; the variance is not split between the two measures). On the possibility that total IQ could be more closely related to problem solving and student performance than all other indications suggested, it was entered first in the second regression analysis.

(Continued)

TABLE 12
First Analysis: Variance Accounted for in Regression Analysis, with the Strongest Variable Allowed to Enter First

(F values significant at .05 for all except Steps 6 and 7* of Unit I)

Variables	Step*	Unit I	Step*	Unit II
Problem Solving	1	37.0%	1	38.7%
Following Directions	2	8.8%	8	.9%
Ability to Use Hands		----	2	7.3%
Absence Ratio (Unit I)	3	4.5%		----
Grouping Ability		----	3	3.4%
Prerequisite Knowledge	4	4.4%	6	1.4%
IQ (Total Score)		----	4	1.7%
Awareness	5	.9%		----
Test Class	6	.6%	7	1.1%
Reason for Placement	7	.6%	5	.9%
Chronological Age		----	9	1.0%
		56.8%		56.4%

*Order in which the variables entered the regression equation.

Differences in Performance Among Classes

Another way of looking at differences in performance is to use the classroom group rather than the individual student as the unit of analysis. Did some test classes learn significantly more or less than others? This is the question studied by covariance analysis.

To accomplish this analysis, the differences in abilities of students in each class at the beginning of the year (or at the start of instruction) must be equalized statistically so that the differences in class means on posttest performance do not merely reflect basic differences in ability of students. Such an adjustment requires a measure of student ability independent of the performance measure. Total IQ scores are traditionally used as the external criterion. In this instance an investigation was conducted to discover whether IQ or problem-solving ability would be the more appropriate measure of entering ability. This question was raised by the finding that the problem-solving measure accounted for the greatest amount of variance in regression analysis. It was hypothesized that problem-solving ability (or combinations including it) would provide the most accurate independent measure of general ability and would produce a higher F value, accentuating the significance of the differences. A second hypothesis was that the actual variance among classes would be narrower if problem solving instead of IQ was used. The subsequent investigation looked into both measures, individually and together, along with problem solving paired with ability to follow directions. Tables 15 and 16 show the results for the four variations of covariates for each unit of instruction.

(Continued)

TABLE 15
Unit I--Analysis of Covariance

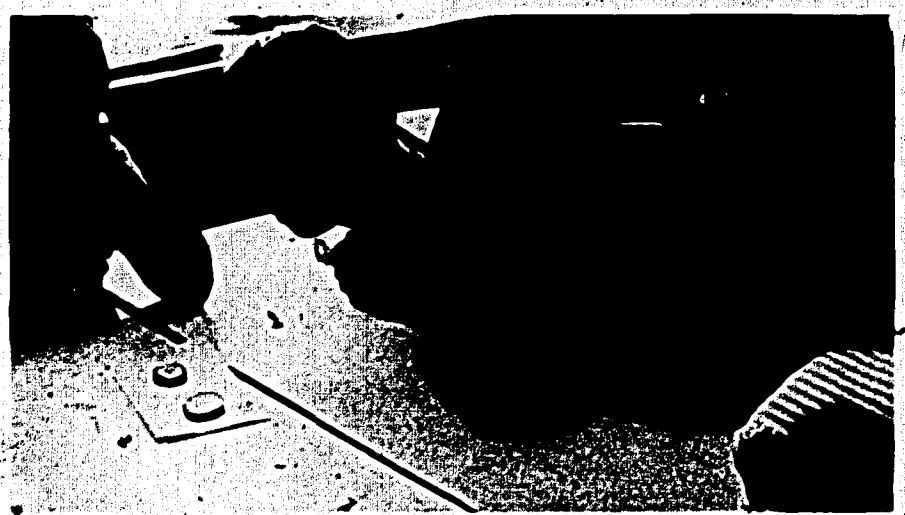
Covariate(s)	F*	Level of Significance		Adjusted Means for Classes (High → Low)									
Total Intelligence (IQ)	2.12	.05	Class ID	06	12	11	00	14	07	05	13	9	8
			Mean	85	78	74	72	71	71	66	61	60	58
Problem Solving (PS)	2.46	.025	Class ID	12	06	11	07	14	00	05	13	9	8
			Mean	81	80	76	73	72	69	66	61	58	58
IQ & PS	2.36	.025	Class ID	06	12	11	14	07	00	05	13	9	8
			Mean	81	80	75	73	72	70	66	60	58	58
PS & Following Directions	2.77	.01	Class ID	06	12	11	07	14	00	05	13	9	8
			Mean	81	80	74	74	71	70	68	60	59	58

*for 9, 98 Degrees of Freedom

TABLE 16
Unit II--Analysis of Covariance

Covariate(s)	F*	Level of Significance		Adjusted Means for Classes (High → Low)									
Total Intelligence (IQ)	4.03	.01	Class ID	06	00	05	11	07	09	13	14	12	08
			Mean	67	54	54	49	49	47	45	44	43	40
Problem Solving (PS)	2.81	.01	Class ID	06	05	07	11	00	12	13	09	14	08
			Mean	62	54	54	51	49	48	45	44	43	41
IQ & PS	3.81	.01	Class ID	06	05	00	11	07	14	12	09	13	08
			Mean	64	53	51	51	50	45	45	45	44	41
PS & Following Directions	3.36	.01	Class ID	06	05	07	11	00	12	13	09	14	08
			Mean	62	55	54	51	50	48	45	44	42	41

*for 8, 98 Degrees of Freedom



Difference in Performance (Continued)

Regardless of the covariate, differences among classes were statistically significant. For Unit I, IQ had the lowest F value. As the regression analysis suggested, a combination of IQ and problem-solving ability did not provide a more significant solution than problem solving alone. A combination of problem solving and ability to follow directions (a teacher rating that added to the explained variance in the regression analysis) provided the highest F value. In line with these findings, support for the second hypothesis was seen, in that for problem solving (and combinations including it) the variance among classes was narrower. This hypothesis was also supported for Unit II; however, the first hypothesis was not. While all F values obtained for Unit II were significant at the .01 level, IQ produced an F value larger than any other set of covariates.

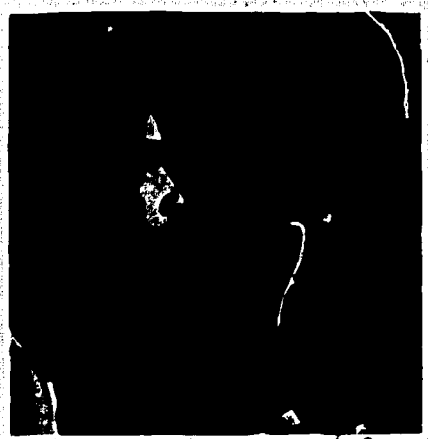
A third consideration in this study of covariates involved the adjusted means for different classes. It was hypothesized that the teachers and classes falling at the extremes would remain the same regardless of the covariates used. This proved to be the case for both units. As the combination of problem solving and ability to follow directions yielded an F value significant at the .01 level for both units, it would seem that these covariates would be most appropriate to use in further analysis of this field test group.

Regardless of the covariate used, one teacher, 06, proved to be significantly more effective in teaching both units of instruction than the other nine teachers. A second teacher, 12, was significantly more effective than the rest in teaching Unit I. On the negative side of the ledger, one teacher, 08, was significantly less effective in teaching both units. Several other teachers, notably 09 and 13, fell well below others in teaching Unit I.

During the field test year, staff members observed the test classes and screened the feedback from teachers. On the basis of these data the staff had identified five teachers who showed considerable fidelity to the strategies suggested in the Teacher's Manual. These teachers, 00, 04, 06, 07, and 12, also appeared to fully understand the intent of the materials. Four of the five classes taught by these teachers yielded complete data for analysis. Two of these four had the highest mean scores, and the other two had mean scores near the high end of the scale.

Several other teachers, 09, 13, 11, and 03, had been rated by the staff as deviating greatly from the intent of the materials. Of these, three had complete data available for analysis and two proved to have mean scores near the low end of the scale; the other class had a lower than average mean score. The staff had rated the teacher whose class had the lowest mean score as showing moderate fidelity to the Manual.

The implication of these findings is obvious. The strategies suggested for teaching the activities appear to be an important factor in student performance.



STUDENT PERFORMANCE ON ITEMS ASSESSING INSTRUCTION IN UNIT 1

After study and validation of the assessment items used during the 1971-72 school year, 65 items were retained for analysis. Twenty one of the retained items assessed the general level of functioning of students in the field test. These items were discussed in a preceding article entitled "Functional Abilities of Students." The remaining 44 items assessed instruction in the 47 activities of Unit I and Part 1 of Unit II (the latter now revised as Unit III).

The original assessment items for each unit were organized into a test booklet and administered by the test teachers immediately prior to, and after, instruction in each unit. Testing occurred in October 1971, February-March 1972, and May-June 1972. Pretests were not administered to four classes participating in a cuing study. (See article entitled "The Development of Test Items.") Of the fourteen test teachers involved in administering the pretests, only one (03) failed to supply the data required for this analysis. Three of the fourteen test classes (01, 03, 04) did not complete the second unit of instruction. Data is therefore reported on thirteen field test classes for Unit I and ten classes for Unit II.

Unit I: Analysis of Performance Items

Twenty five of the 44 items directly related to instruction occurred in Unit I. (The other 19, for Unit II, will be discussed in Article 8.) The 25 items assessed instruction presented in the 26 activities of the unit, and the items fell into four categories by topic: 1) directionality and map reading; 2) measurement and scale; 3) temperature; and 4) environmental comparisons. Although the results are presented by these content topics, the items were interpreted separately rather than as only "subtest scores." The statistical analysis reported in the preceding articles made use of only weighted total scores for the two unit tests.

Pre- and posttest scores for each item will be shown by subtest, and while it is possible to discuss the results of all of these test items in terms of gains in instruction, it seems more appropriate at this point in field testing to give greater attention to posttest, rather than pretest, performance. Therefore, unless the percentage of students responding correctly was essentially the same on both pre- and posttests, reference will be made only to the posttest percentages.

The percentage of correct response on the earlier ME NOW test items was characteristically at the 50 percent level. In that study many students had reasonable and logical explanations for choosing options other than the intended answers to many items; hence, the level of understanding was not accurately reflected in the percentage of students choosing the "right" answer. A similar condition also has occurred in the ME AND MY ENVIRONMENT assessment. In fact, 40 percent of the Unit I test items have a qualitative scoring key, with some responses given credit as partially correct. The summary table of responses by class, however, reflects *only* the percentage of students choosing the most acceptable answer for each item. The reasons for choosing other "answers" provide a separate justification for inspecting each item separately.

Interpretation of Results

The problem of assessing the learning of EMH students is not resolved by producing an "achievement test," since the question of what standard to apply to performance on these items or subtests is one that is difficult to answer. It is unrealistic to expect all children to be able to answer all items, when a wide range of difficulty and a range of topics are involved. Answering even one or two more items correctly on the posttest than on the pretest may represent considerable learning for some students. The items represent the staff's judgment of key content that should be learned, *not* all learning that can be expected to occur in instruction. Some areas of learning were not assessed at all; others were explored only through interviews and are not reported here. Notably absent are measures of observation skills and problem-solving skills that are directly related to the materials. Students' attitudes also are not reflected here, although measures are reported elsewhere. The effects of this curriculum on self-confidence, social participation, task orientation, and general response to school have not been assessed.

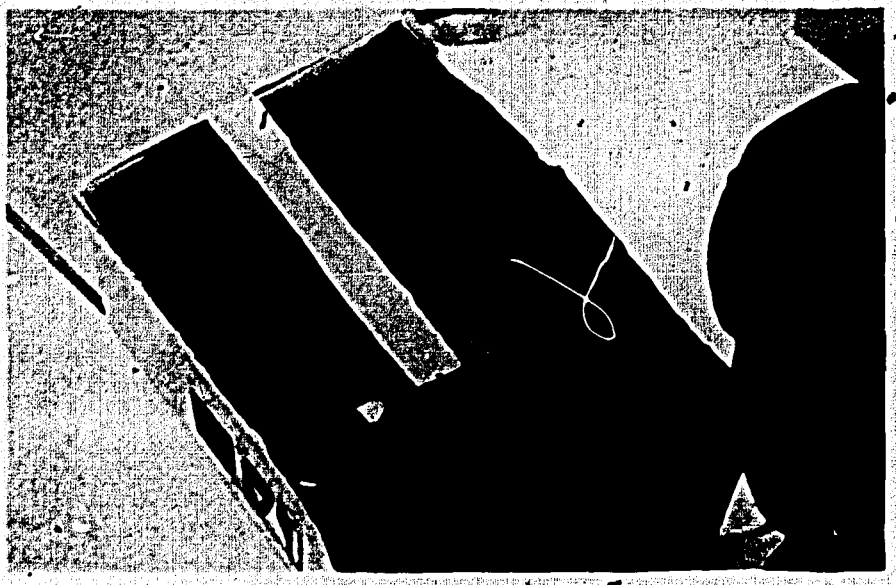
Efforts were made from the beginning to assure that use of the items in different classes would be comparable. The tests were administered by the teachers, using the same instructions. Some training was provided at the beginning of the school year, and additional written instructions were sent for each of the four testing periods. Even with these precautions, however, conditions and procedures were not standardized. (For example, total test time for the pretest on Unit I ranged from 30 minutes to 85 minutes and was distributed over one to three days.)

The test items themselves were undergoing their first field test. Individual interviews with students were conducted to validate some of the items, and, as a result, approximately 20 percent of the original number of items were eliminated as defective. Undiscovered weaknesses may remain in some of the 25 items retained for analysis.

Only a small number of items originally existed for each instructional area to be assessed. The 20 percent loss of defective items contributed to an imbalance in the areas actually assessed. Although some broad understandings were assessed by interviews, teacher ratings, and observations, and are reported elsewhere, the results and conclusions both there and here should be interpreted with caution.

With all these qualifications considered, the test items do represent the best judgment of the staff as to key concepts to be learned in the materials. Even though the results should not be used as summative evaluation of student learning, they do provide a clue to the degree of learning occurring. Combined with the information on the functional abilities of this population, expectations of student response can be readjusted, and revisions designed to further enhance learning.

A summary of results from the use of the test items follows.



Unit I, Subtest I, Directionality and Map Reading

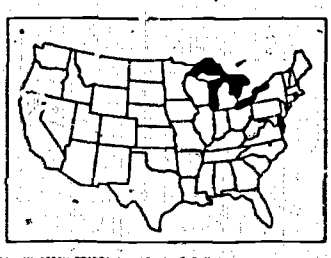
Table 17 shows a summary of Subtest 1 statistics. Individual items, their scoring keys, and the percentage of students in the total group who selected each option on the pre- and posttest are shown. The percentage of students, by class, who chose the "best" response to each item is shown in Table 18.

These items had a wide range of difficulty, from Item 25, which almost none of the students answered correctly, to Item 2, which about two thirds answered correctly. About one third of the students in 4 of the 13 classes were able to answer most of the items in this subtest; two thirds or more of the students in two classes, fewer than two items.

Of the five items related to directionality and map reading, two (Items 4 and 6) involved a student's ability to orient himself in space. These two items required a knowledge of which side of a map is north, a knowledge of compass directions, and orientation for all directions when presented on a map. On the posttest, roughly two thirds of the total group placed north at the top of the map. When given north, half of the total group correctly labeled the other three compass directions.

2.

THIS IS A MAP OF THE UNITED STATES.



DRAW AN ARROW POINTING NORTH ON THE MAP.
MARK AN X AT THE PLACE WHERE YOU LIVE (WHERE YOUR TOWN IS.)

	Pre	Post
1 = North	75	88
2 = South	66	66
3 = East	156	65
4 = West	35	106
5 = Corner	75	75
6 = Multiple Response	75	16
7 = No Response	138	116

TABLE 17
Summary Information for Subtest 1
Directionality and Map Reading
(five items)

This group of items measured:
--knowledge of which side of a map is north (Item 2)
--knowledge of compass directions and orientation of direction on a map (Items 4, 6)
--ability to locate the intersection of two streets on a map (Item 25)
--ability to determine a reasonable route to take between two places on a map (Item 26)

Subtest Weighted Scores:

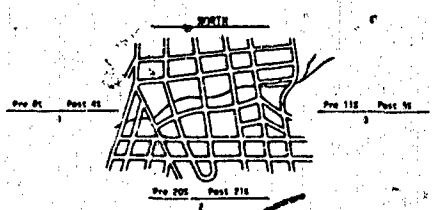
Item	Weights Per Option 1	2	3	4	5	6	7
2	5	0	0	0	0	0	0
4	3	3	3	5	0	4	0
6	0	5	0	0	5	0	0
25	5	4	3	2	0	0	0
26	5	4	2	0	2	0	0

Subtest Statistics:

Maximum possible score: 25
Range of Scores: 0-24
Mean: 11
Standard Deviation: 6.7

4.

THE MAP BELOW SHOWS PART OF A CITY. NORTH IS MARKED ON A LINE AT THE TOP OF THE MAP.
MARK EAST, WEST, AND SOUTH, IN THEIR RIGHT PLACES, ON THE OTHER THREE LINES.



	Pre	Post
0 = all	69	68
1 = none	225	141
6 = any two correct	16	0
7 = no response	45	89

Two items in Subtest 1 dealt with locating things on a map. One (Item 25) asked students to trace a street route from one landmark to another. Fifty percent of the students did so on the posttest. However, for another item (Item 6) only one third of the students correctly indicated what direction one landmark was from another – the same percentage as on the pretest. The most difficult item in this subtest dealt with the ability to locate the intersection of two streets on a map. This item required that the student be able to recognize that a signpost on a street corner provides the information of what *two* oppositely oriented streets to find and trace to their intersection on the map. Only six students out of 172 (or three percent) correctly did this on the posttest, and all but one of these six students were in the same classroom.

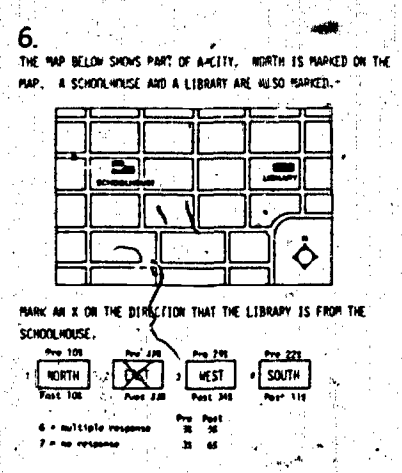
A random sample of students was interviewed about Items 2, 4, 25, and 26 in this subtest. For Item 2, typical responses to the interview were to point to the top of the map to indicate north. For Item 4, identifying east, west, and south on the map was more difficult; however, the intent of this question, too, was clear to the students. For Item 25, locating an intersection on a map involved numerous problems. Many students were not able to derive the names of the streets from the street sign in the picture of Bill's house. Some simply placed a mark somewhere on the map directly below the street corner in the picture. Some tried to track the direction of the street in the picture directly down onto the map. These related but incorrect efforts probably accounted for many of the responses coded 4 and 5 on this item. The inability of many students to find the two different streets around the margins of the map and trace them to their point of intersection was a major problem identified by student interviews. Even when given the streets and their locations on the map, a number of students were unable to follow the two coordinates to the point at

(Continued)

which they crossed. Having marked a location for Bill's house, many students understood, in Item 26, how to start from that point and draw a line showing where he went on his bicycle. Some students were confused by the art work and labeling of the streets on the map. Still other students simply drew a line indicating Bill going any place on his bicycle and did not understand that they were to find a path from home to school.

Table 18 indicates the percentage of students in each class whose total subtest score was more than one standard deviation above or below the total group mean. This is essentially a table showing the percentage of students in each class who were high performers, answering more than three items correctly, or low performers, answering fewer than two items correctly. (Note that partial credit was possible on three of the five items.)

Considering Items 2, 4, 6, 25, and 26 as a subtest, 21 percent of the students in the total group answered fewer than two items, and only 20 percent answered more than three items, indicating a generally very low level of performance in directionality and map reading after instruction (mean: 11, S.D.: 7, maximum score possible: 25). Based on these results, additional activities were written into the materials to provide practice in recognizing and using compass directions and in making use of a map to trace routes from one landmark to another. Investigation of Item 25 on locating an exact point at an intersection suggested that students had difficulty using coordinates in *any* way; therefore, an activity was written dealing with graphing, and practice was provided in locating a number of coordinates, with the teacher's attention drawn to using two variables and locating intersections.



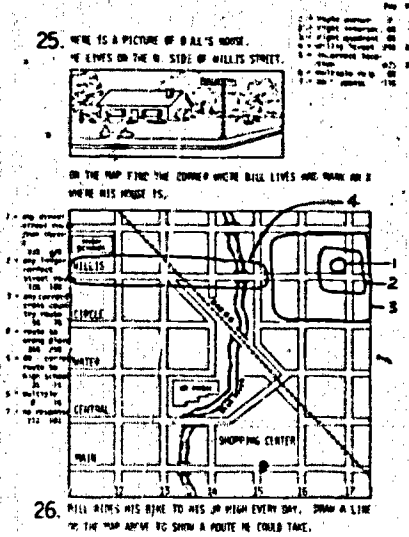


TABLE 18
Subtest 1 (Directionality & Map Reading): Percent Choosing the Correct Response, by Class

	Field Test Classes																								All Classes Combined	
	00		01		02		04		05		06		07		08		09		11		12		13		14			
Question	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post		Post		Post		Post		Post	Pre	Post
N	16	14	18	9	18	16	9	9	12	4	9	9	17	10	12	12	13	11	9		15		11		19		124	148
2	69	50	31	69	43	47	36	44	20	36	44	67	54	77	40	42	75	67	79		88		57		80		47	53
4	23	57	23	77	21	42	7	22	40	55	56	78	69	47	40	42	50	58	64		71		50		60		34	53
6	15	6	23	38	36	35	43	22	30	36	33	56	31	29	50	42	50	25	50		35		36		36		33	33
25	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	42	0		0		0		0		0	3
26	31	50	8	0	29	35	29	0	10	36	44	44	23	59	30	33	25	25	57		41		57		36		35	50
Level of Performance	Posttest only Subtest Performance*																											
High (more than three items correct; score of 18 or more)	0		0		0		0		0		33		40		25		36		22		27		45		26		20	
Low (fewer than two items correct; score of 4 or less)	29		89		56		11		0		0		10		11		0		11		7		9		16		21	

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the

Unit I, Subtest 2, Measurement and Scale

Table 19 shows a summary of Subtest 2 statistics. Individual items and their scoring keys are also shown, along with the percentage of students in the total group who selected each option on the pre- and posttest. Table 20 shows the percentage of students, by class, who chose the best response to each item. The general level of performance on these items ranged from one fifth to one half of the students who answered an item correctly.

The six items in this subtest assessed ability to estimate distances and an understanding of the use of scale to determine distance on a map. Two of the items (Items 7 and 41) dealt with actually estimating distance, which required students to have at least a rough approximation of the meaning of units of length, such as feet. Less than 20 percent of the total group made reasonable estimates after instruction, although almost twice this many (39 percent) indicated that pacing a distance was a procedure that could be used to make such estimates. Three of the items (Items 12, 16, and 23) dealt with using a scale given on a map or drawing to determine length or distance. Over one third of the total student group proved successful on these items on the posttest.

A random sample of students was interviewed regarding their understanding of Items 7, 12, and 16. On Item 7, almost two thirds of the students interviewed could correctly explain that a pace was a step. Only one in ten students, however, knew how long his pace was. Less than half of those who marked the correct response answered correctly what a pace was or how long

(Continued)

TABLE 19

Summary Information for Subtest 2 Measurement and Scale (six items)

This group of items measured:
- --understanding of the use of pacing to estimate distance (Items 7 and 42)
- --ability to make a reasonable estimate of distance (Item 41)
- --ability to determine length using scale drawings (Items 12 and 23)
- --ability to use the scale to determine distance on a map (Item 16)

Subtest Weighted Scores:

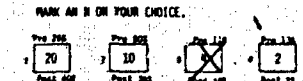
Item	Weights Per Option 1	2	3	4	5	6	7
7	0	0	5	0	-	0	0
12	5	0	0	0	-	0	0
16	0	5	0	0	-	0	0
23	0	5	2	0	-	0	0
41	0	3	5	-	-	0	0
42	5	0	0	0	0	-	-

Subtest Statistics:

Maximum possible score: 30
Range of Scores: 0-28
Mean: 10.2
Standard Deviation: 6.6

7.

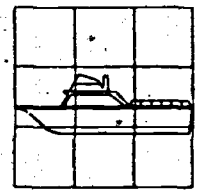
ABOUT HOW MANY PACES ARE THERE IN TEN FEET?
MARK AN X ON YOUR CHOICE.



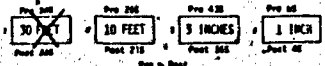
	Pre	Post
6 = multiple response	0	1%
7 = no response	0	1%

12.

EACH SQUARE IN THE GRID BELOW IS ONE INCH ON A SIDE.
WHAT IS THE LENGTH OF THE REAL OBJECT PICTURED ON THE GRID: 30 FEET, 10 FEET, 5 INCHES, 1 INCH



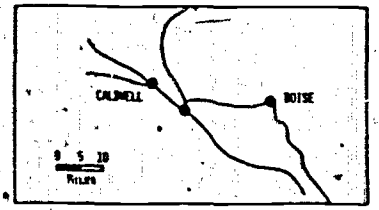
MARK AN X ON YOUR CHOICE.



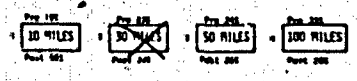
	Pre	Post
6 = multiple response	1%	1%
7 = no response	3%	1%

16.

BELOW IS A PORTION OF AN IDAHO MAP.



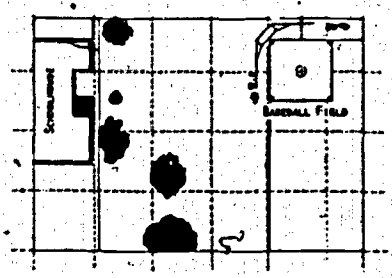
HOW FAR IS IT BETWEEN CALDWELL AND BOISE?
MARK AN X ON YOUR CHOICE.



	Pre	Post
3 = multiple response	2	1%
7 = no response	3%	3%

23.

THE MAP BELOW SHOWS PART OF A CITY. EACH SQUARE ON THE MAP EQUALS 100 FEET.



MARK AN X ON HOW FAR IT IS FROM THE SCHOOLHOUSE TO THE BASEBALL FIELD.

	Pre	Post
6 = multiple response	1%	1%
7 = no response	3%	1%

41. & 42. GUESSING GAME (ESTIMATING DISTANCE)

(The following task was administered to students individually by the teacher who recorded their estimate and method of estimation. Teachers were instructed to place the pieces of tape fifteen feet apart.)

Directions to students: There are two pieces of tape on the floor outside the room. We want to find out who can guess closest to how far it is between the two pieces of tape. The only catch to it is that you can't use a ruler. I'll ask each of you one at a time to try it while the rest of you are drawing a picture. After everyone has a chance to guess, we'll put the answers on the board and see who is closest.

Method of Recording:

ID	Name	Estimate	Method: Pacing	Guess	Other (comment)
—	———	———	———	———	———
—	———	———	———	———	———
⋮	⋮	⋮	⋮	⋮	⋮

41. STUDENT ESTIMATE

Post Only

26%	1 = Way under or way over (3 units or more above or below correct answer
24%	2 = Approximate (plus or minus up to 3 units)
9%	3 = Correct answer
25%	4 = Inappropriate or wrong response
5%	7 = No response

42. METHOD OF ESTIMATION

Post Only

39%	1 = Pacing
10%	2 = Guess
2%	3 = Instrumental use (floor tiles or something, etc.)
24%	4 = Other
5%	7 = No response

one was. For Item 12, the interviews revealed that many students had no notion of how to use the scale provided. A number of those who marked the correct option said they just guessed or thought that the real boat could be thirty feet long. Some answered the question only in terms of the pictured drawing and appeared to have no concept of it as a scale drawing.

In the interviews, about two thirds of the students could actually trace on Item 16 the route from Caldwell to Boise with a pencil. Of those who had marked the correct response, only half understood the use of the scale for miles; the others guessed. A significant finding was that one student in three did not even understand the map symbolism which indicated where the two cities were and the route between them. Some traced beyond the two dots signifying the cities; others did not follow the road.

Table 20 indicates the percentage of students in each class whose total subtest score was more than one standard deviation above or below the total group mean. High performers on this basis obtained a score of 17 or more out of 30 points, or more than three of the six items correct. Low performers obtained less than three points, which meant answering less than one complete item correctly. (Partial credit was possible on two of the six items.)

Considering the six items as a subtest, 20 percent of the students answered more than three of the six items correctly, while 18 percent answered none correctly. Thus the level of performance on measurement and scale was also quite low after instruction. The materials have been revised to provide a great deal more practice in estimating and measuring very short distances. Revised activities begin at a much more basic level on the use of a ruler, and students are given many more opportunities to measure short lengths with it. The concept of scale, and problems dealing with scale, have been eliminated entirely from revised materials.

TABLE 20
Subtest 2 (Measurement & Scale): Percent Choosing the Correct Response, by Class

Field Test Classes

Question / N	00 Pre 16	00 Post 14	01 Pre 18	01 Post 9	02 Pre 18	02 Post 16	04 Pre 9	04 Post 9	05 Pre 12	05 Post 4	06 Pre 9	06 Post 9	07 Pre 17	07 Post 10	08 Pre 12	08 Post 12	09 Pre 13	09 Post 11	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined Pre 124	All Classes Combined Post 148
7	0	25	31	15	7	12	7	22	0	27	11	0	8	6	10	0	17	50	21	24	21	16	11	18
12	15	25	0	8	14	29	21	22	30	9	0	22	23	59	40	33	25	42	29	47	29	52	20	35
16	31	31	8	15	36	12	14	44	30	9	33	67	23	41	30	42	8	42	36	41	50	20	22	34
23	23	69	15	46	71	71	50	78	60	36	44	56	54	47	40	58	42	58	57	76	29	40	43	55
41		25		0		0		0		9		44		6		8		0	0	18	14	0		9
42		6		31		53		67		18		89		18		100		92	21	59	21	24		39

Level of Performance — Posttest only Subtest Performance*

High (more than three items correct; score of 17 or more)	7	11	0	0	0	56	10	22	55	11	33	36	5	20
Low (less than one item correct; score less than 3)	7	33	25	56	50	0	10	0	0	11	13	27	21	18

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the mean.

Unit I, Subtest 3, Temperature

See Table 21 for a summary of Subtest 3 statistics, the four items, keys, and pre- and posttest results. Class-by-class results are shown in Table 22. Considering the general level of performance on individual items, three fourths or more of the students answered three of these four items correctly before instruction. Except for Item 20 little change in performance was noted on the posttest.

One item (Item 3) assessed the students' understanding of how to read a thermometer. It was answered correctly by three fourths of the students on the posttest, indicating no gain from pretest responses. A second item (Item 22) assessed awareness of where temperatures are likely to be highest outdoors. Again, three fourths of the students responded correctly on the posttest, and again this represented no gain from pretest responses. A third item (Item 11) assessed knowledge that dark colors absorb more heat than light colors. About two thirds of the students answered this item correctly on the posttest. The fourth item (Item 20) assessed recognition of a record of outside temperatures expressed as a graph and compared with a graph of indoor temperatures. Almost four fifths of the students answered this item correctly on the posttest.

A random sample of students was interviewed about their understanding of Items 3, 11, and 20. Almost 40 percent of the students interviewed could correctly identify the Fahrenheit scale given in Item 3, as well as give a reasonable explanation for answering the item as they did. In one out of five cases, the students acknowledged that they simply guessed on the item. Of those who marked the correct response for Item 11, 40 percent did not have any idea of the concept that white reflects heat, nor any memory of a science activity related to this idea. Thus, the results

(Continued)

TABLE 21
Summary Information for Subtest 3
Temperature
(four items)

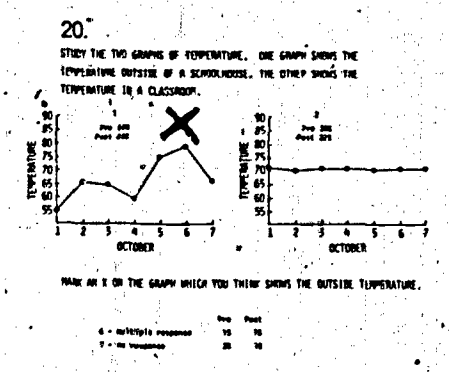
This group of items measured:
--understanding of how to read a thermometer (Item 3)
--awareness of where outdoor temperatures are likely to be highest (Item 22)
--knowledge that dark colors absorb more heat than light colors (Item 11)
--recognition of a record of outside temperatures expressed as a graph (Item 20)

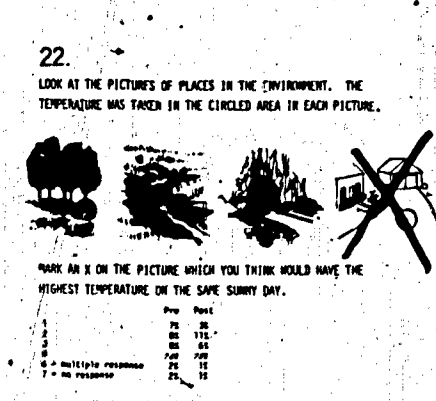
Subtest Weighted Scores:

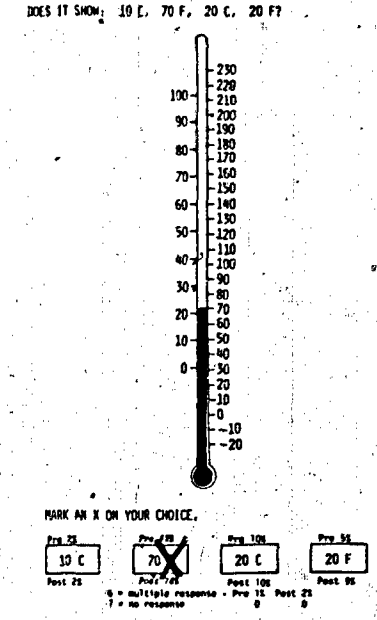
Item	Weights Per Option 1	2	3	4	5	6	7
3	0	5	0	0	-	0	0
11	0	5	-	-	-	0	0
20	5	0	-	-	-	0	0
22	0	0	0	5	-	0	0

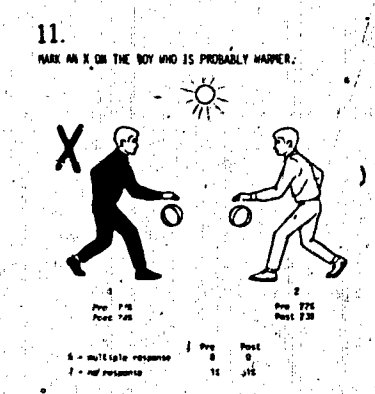
Subtest Statistics:

Maximum possible score: 20
Range of Scores: 0-20
Mean: 11.7
Standard Deviation: 7









are an overestimate of student understanding of this concept. Item 20 involved a minimum of guessing on the students' parts. They responded on the basis that outside temperatures vary while inside temperatures remain the same. Other than noticing this difference in the shapes of the graphs, most students did not understand the graphs and could not read them. Many could not visually trace the points on the graph to its scales for the temperature or the day.

When this group of items was considered as a subtest, the average was determined as answering two of the four items correctly. Twenty two percent of the students answered all four items correctly; 17 percent answered none correctly (mean: 12, S.D.: 7, maximum score possible: 20). Table 22 shows the percentage of high and low performers, by class. Compared to performance on the other subtests, students were slightly more successful with this group of items. Considering pretest levels for this subtest, however, gains from instruction were disappointing. Several changes were made in the related activities during revision. Where there was one activity involving temperature in the original materials, there are now four activities. One of the new activities provides practice in reading a thermometer, while two more provide many opportunities for students to use the thermometers to measure temperatures. The fourth new activity on temperature has been developed on reading and making graphs, providing much practice in this skill.

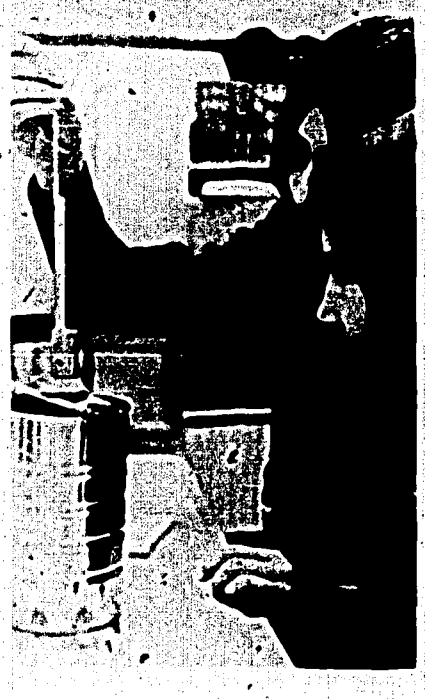


TABLE 22

Subtest 3 (Temperature): Percent Choosing the Correct Response, by Class

Field Test Classes

Question N	00 Pre 16	00 Post 14	01 Pre 18	01 Post 9	02 Pre 18	02 Post 16	04 Pre 9	04 Post 9	05 Pre 12	05 Post 4	06 Pre 9	06 Post 9	07 Pre 17	07 Post 10	08 Pre 12	08 Post 12	09 Pre 13	09 Post 11	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined: Pre 124	All Classes Combined: Post 148
3	69	69	92	85	100	82	93	89	80	55	78	78	92	88	50	83	75	67	93	82	45	84	82	76
11	85	44	54	69	57	65	71	89	90	82	89	89	77	88	70	83	92	83	71	100	73	64	77	74
20	38	56	31	23	57	65	71	89	40	55	78	89	77	65	60	67	58	58	50	76	64	68	58	65
22	92	81	62	77	71	71	93	89	60	82	78	67	54	88	90	100	83	67	93	65	86	72	74	78

Level of Performance — Posttest only Subtest Performance*

High (all correct; score of 20)	14	0	0	0	0	44	30	33	18	44	40	18	32	22	
Low (none correct; score of 0)	0	78	56	89	0	0	0	0	0	11	0	0	0	37	

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the mean.

Unit I, Subtest 4, Environmental Comparisons

Table 23 contains a summary of Subtest 4 statistics, items, keys, and pre-posttest results. Class-by-class results are shown in Table 24. Ten items were included in this subtest, and five allowed partial credit for some responses. Many of the ten items had a low level of correct response on the pretest and showed marked gains in correct response on the posttest. From half to three fourths of the students made correct responses to individual items on the posttest.

TABLE 23
Summary Information for Subtest 4
Environmental Comparisons
(ten items)

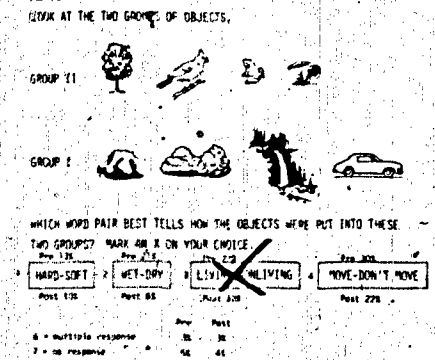
- This group of items assessed:
  - --the awareness of how things are affected or changed in an experiment (Items 8, 9, 10)
  - --the recognition of concrete or intangible, near-at-hand or remote features of the environment (Item 13)
  - --an understanding of the categories living-nonliving (Items 19, 24, 28)
  - --recognition of nonessentials versus necessities for life (Item 27)
  - --knowledge of human and plant needs for life (Items 32 and 33)

Subtest Weighted Scores:

Item	Weights Per Option 1	2	3	4	5	6	7
8	0	5	0	0	-	0	0
9	5	0	0	0	-	0	0
10	0	0	0	5	-	0	0
13	0	1	3	3	5	5	0
19	0	0	5	0	-	0	0
24	0	0	5	0	-	0	0
27	0	3	5	5	5	0	0
28	5	3	0	0	0	0	0
32	5	5	4	3	0	0	0
33	5	5	5	4	3	0	0

Subtest statistics:

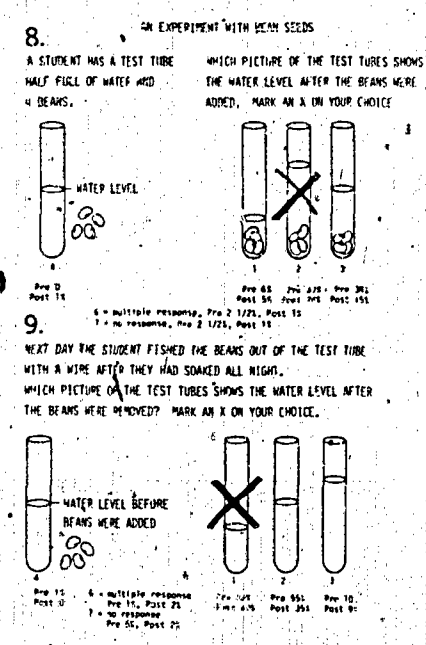
Maximum possible score: 50
Range of Scores: 3-49
Mean: 29.4
Standard Deviation: 11.7

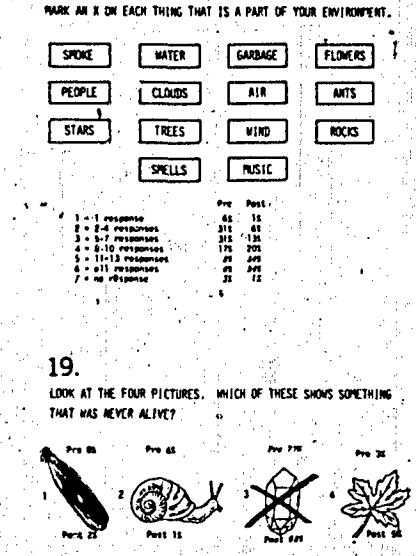


Three items (Items 8, 9, and 10) were concerned with awareness of how things are affected or changed in an experiment which the students had performed. (The information assessed had not been called for by the experiment, however.) On the posttest, from one half to three fourths of the students marked the appropriate responses. For many classes this represented a considerable increase in correct responses. An item requiring multiple responses (Item 13) assessed the recognition of concrete or intangible, near-at-hand or remote features as components of the environment. Fifty eight percent of the students identified most or all of the items as components of the environment on the posttest, a dramatic increase over pretest performance. Three items (Items 19, 24, and 28) assessed an understanding of the categories living/nonliving. Two of these required simple recognition of the categories, and on these items, from two thirds to four fifths of the students responded correctly. However, only half of the students could supply the label living/nonliving to a set of items which required this description.

One item (Item 27) assessed recognition of nonessentials as opposed to necessities for life. Thirty nine percent of the students responded correctly to this item, three times the pretest percentage of correct responses. Finally, two items (Items 32 and 33) required written responses listing human and

(Continued)





Environmental Comparisons (Continued)

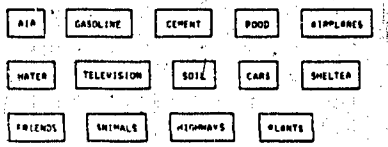
plant needs for life. From one half to two thirds of the students responded correctly on the posttest.

Table 24 shows a summary of high and low performers on this subtest. Considering all ten items as a subtest, 17 percent of the students in the total group answered eight or more of the 10 items, while 20 percent answered fewer than four of the items (mean: 28, S.D.: 12, maximum score possible: 50). The level of posttest performance on this subtest was relatively high, with the average determined as answering more than half of the items correctly. Individual classes demonstrated considerable gains on various items as well. To remedy some weaknesses, the revised materials were given a great many more directions and examples for categorizing, observing, and comparing things. Revisions were made in the sequence of instruction related to the concept of living/nonliving, and changes were made in the treatment of plant and animal needs for life.

In summary, the performance on 25 measured items related to Unit I of ME AND MY ENVIRONMENT was low. There were evidences of learning on some items, as indicated by gains from pre- to posttest scores, but the overall results do not reflect levels of learning considered acceptable by the project staff. An increase in the number of students understanding the key concepts is sought in the extensive revision of Unit I.

27.

MARK AN X ON EACH THING WHICH YOU COULD LIVE WITHOUT.



28.

MARK AN X ON EACH PICTURE OF A LIVING THING.

32.

NAME _____

WRITE DOWN THE THINGS THAT YOU MUST HAVE TO STAY ALIVE.

33.

NAME _____

WRITE DOWN THE THINGS THAT PLANTS MUST HAVE TO STAY ALIVE.

TABLE 24

Subtest 4 (Environmental Comparisons): Percent Choosing the Correct Response, by Class

Field Test Classes

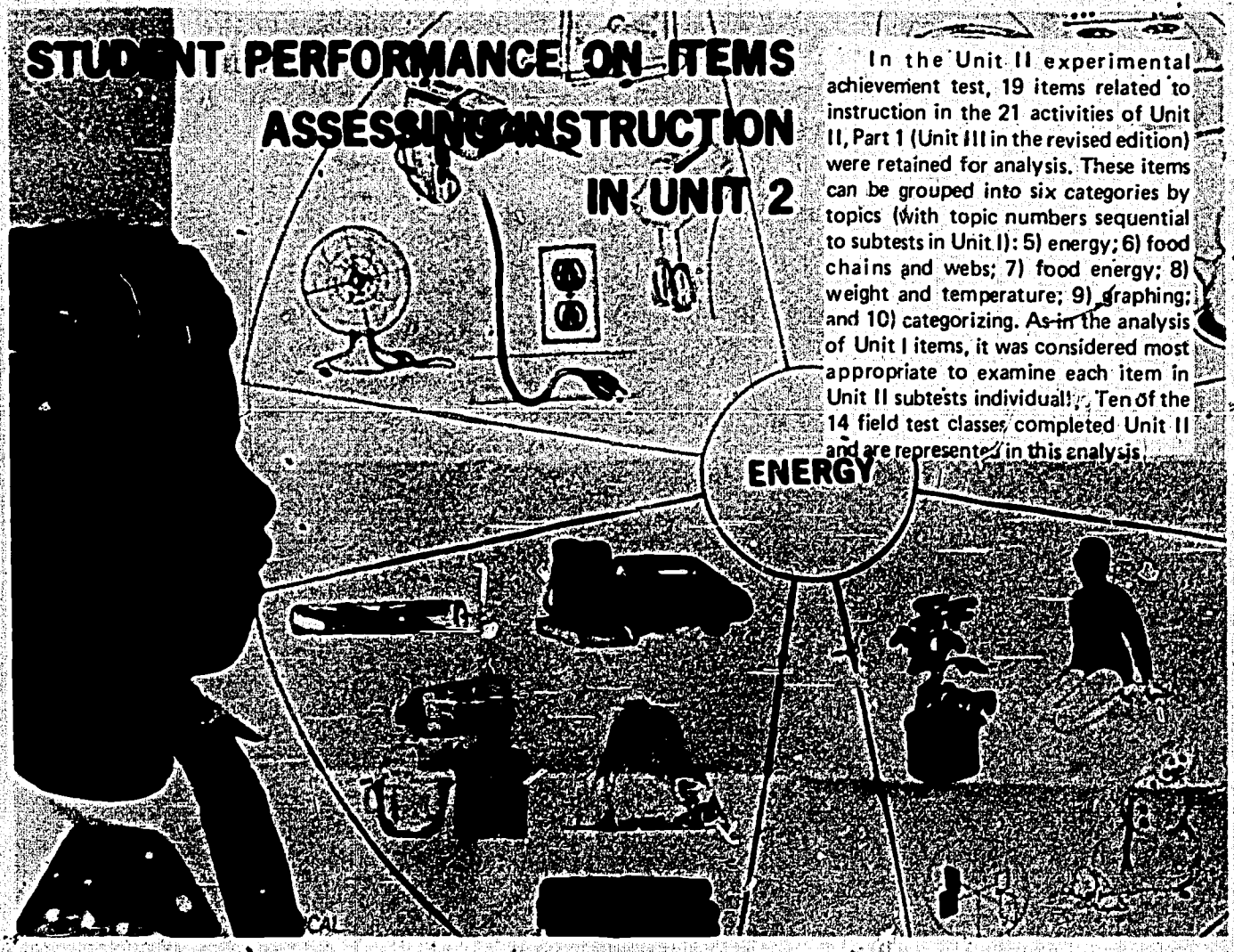
Question N	00 Pre 16	00 Post 14	01 Pre 18	01 Post 9	02 Pre 18	02 Post 16	04 Pre 9	04 Post 9	05 Pre 12	05 Post 4	06 Pre 9	06 Post 9	07 Pre 17	07 Post 10	08 Pre 12	08 Post 12	09 Pre 13	09 Post 11	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined Pre 124	All Classes Combined Post 148
8	69	75	38	31	50	82	57	78	60	82	56	89	62	71	20	67	50	92	64	88	86	64	51	76
9	15	50	23	62	21	59	36	44	30	45	33	78	46	35	60	58	17	50	71	59	43	36	29	50
10	23	69	23	23	57	59	50	67	50	64	44	56	38	53	30	42	50	58	64	41	54	40	39	53
13	15	19	0	0	0	18	0	44	0	9	11	11	0	6	10	33	8	42	50	18	14	36	14	24
19	69	75	54	77	79	88	93	100	80	100	78	100	92	94	60	75	75	75	93	100	86	88	77	88
24	23	63	8	54	36	53	43	56	20	45	11	78	46	35	20	42	33	50	57	71	64	40	27	52
27	0	31	0	23	14	53	14	66	0	18	11	55	46	47	10	17	16	42	21	42	42	40	13	39
28	62	56	38	38	57	0	64	89	40	9	78	78	46	88	70	67	75	17	43	77	78	72	59	70
32	69	62	8	23	71	82	64	67	80	73	44	89	30	82	30	66	25	75	79	41	64	64	48	69
33	23	38	0	15	36	65	21	44	50	63	44	55	31	59	20	50	25	67	64	18	43	40	26	49

Level of Performance — Posttest only Subtest Performance:*

High (eight or more items correct; score of 41 or more)	43	0	0	0	0	44	30	11	45	0	27	18	0	17
Low (fewer than four items correct; score of 16 or less)	7	89	56	56	0	0	0	0	9	11	0	9	11	4	20

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the mean.

STUDENT PERFORMANCE ON ITEMS ASSESSING INSTRUCTION IN UNIT 2



In the Unit II experimental achievement test, 19 items related to instruction in the 21 activities of Unit II, Part 1 (Unit III in the revised edition) were retained for analysis. These items can be grouped into six categories by topics (with topic numbers sequential to subtests in Unit I): 5) energy; 6) food chains and webs; 7) food energy; 8) weight and temperature; 9) graphing; and 10) categorizing. As in the analysis of Unit I items, it was considered most appropriate to examine each item in Unit II subtests individually. Ten of the 14 field test classes completed Unit II and are represented in this analysis.

Unit II, Subtest 5, Energy

Table 25 shows a summary of Subtest 5 statistics followed by individual items and their scoring keys, and the percentage of students in the total group who selected each option on the pre- and posttest. Table 26 shows the percentage of students by class who chose the best response to each item.

Of the seven items related to energy, two items (Items 18 and 19) assessed understanding that energy is required to do work. These two items proved deceptively easy in that most students could answer them prior to instruction. Posttest levels of correct response were 91 and 85 percent.

(Continued)

TABLE 25
Summary Information for Subtest 5
Energy
(seven items)

This group of items assessed:
--understanding that energy is the ability to do work (Items 18 and 19)
--awareness that the higher the temperature of a bend, the more work it can do (Item 21)
--recognition of things containing stored energy (Item 28)
--knowledge that living things ultimately get their energy from the sun (Item 12)
--knowledge that food is the body's source of energy (Items 9 and 31)

Subtest Weighted Scores:

Item	Weights Per Option 1	2	3	4	5	6	7
9	0	0	0	5	-	0	0
12	0	5	0	0	-	0	0
18	0	0	0	5	-	0	0
19	0	0	0	5	-	0	0
21	0	0	0	0	5	0	0
28	0	1	3	5	1	0	0
31	0	5	0	0	0	0	0

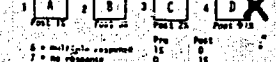
Subtest Statistics:

Maximum possible score: 25
Range of Scores: 0-24
Mean: 18.5
Standard Deviation: 7.2

18.

A TABLET OF EASTER EGG DYE WAS DROPPED AT THE SAME TIME IN EACH OF THE FOUR JARS SHOWN ON THE SLIDE AND LEFT ALONE FOR TEN MINUTES. EACH JAR HAS THE SAME AMOUNT OF WATER.

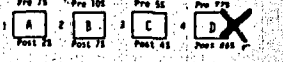
IN WHICH JAR WAS THE MOST WORK DONE?
MARK AN X ON YOUR CHOICE.



19.

IN WHICH JAR DID THE WATER HAVE THE MOST ENERGY?

MARK AN X ON YOUR CHOICE.



Another item (Item 21) assessed the awareness that the higher the temperature of a liquid, the more work it can do. Only 40 percent of the students knew this on the posttest.

Two items proved quite difficult for students. One of them (Item 28) assessed the recognition of things containing stored energy. Almost none of the students answered this item successfully. The other (Item 12) assessed the knowledge that living things ultimately get their energy from the sun. Less than one fifth of the students were able to answer this item correctly on the posttest, representing no change from responses on the pretest. Neither of these items assessed specific points of learning from direct instruction. Instead they required students to see the implications or make inferences about what had been presented in class. In the revision both of these concepts were judged to be peripheral to the major thrust of the unit, and no attempt was made to devise activities to teach this information.

The remaining two items in this subtest (Items 9 and 31) assessed the knowledge that food is the body's source of energy. From half to two thirds of the students were able to answer these items correctly on the posttest, representing considerable gains from pretest scores.

Table 26 indicates the percentage of students in each class whose total subtest score was more than one standard deviation above or below the total group mean. This table shows the percentage of high performers (answering more than five items correctly) and low performers (answering fewer than three items correctly). Note that it was possible to earn partial credit on Item 28.

Considering all of these items as a subtest, 14 percent of the students answered more than five items correctly on the posttest, while 20 percent answered fewer than three items correctly (mean: 18.5, S.D.: 7, maximum score possible: 35).

In only one class were one third or more of the students able to answer more than five items correctly. However, if one discounts the two quite difficult items which were not a part of direct instruction, the average performance on this subtest was at the relatively high level of more than three of the five remaining items answered correctly. Revisions of the materials included breaking activities into smaller steps, experience with more forms of energy, and some practical applications of the idea that the higher the temperature, the more energy is present. The section of activities relating to food as the body's source of energy was expanded with additional activities on the energy values of various foods. A game was developed called "The Full and Healthy Game" dealing with balanced diets and daily food and energy requirements.

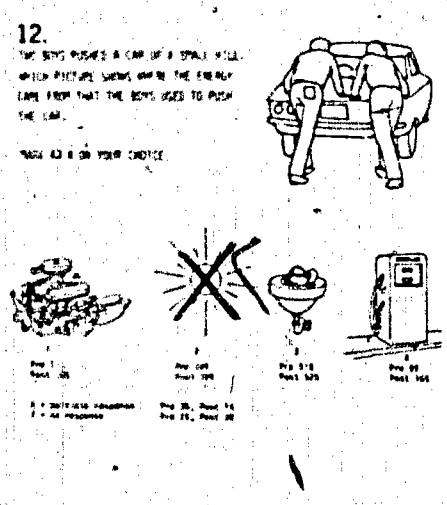


TABLE 26
Subtest 5 (Energy): Percent Choosing the Correct Response, by Class

Question	Field Test Classes 00 Pre	00 Post	05 Pre	05 Post	06 Pre	06 Post	07 Pre	07 Post	08 Pre	08 Post	09 Pre	09 Post	11 Post	12 Post	13 Post	14 Post	All Classes Combined Pre	All Classes Combined Post
N	16	14	10	4	10	9	17	10	12	9	12	11	9	15	11	19	77	111
9	25	53	20	67	10	60	47	53	33	40	67	55	46	60	39	44	32	51
12	6	47	40	0	30	30	6	7	25	30	25	9	0	20	31	9	19	19
18	81	93	100	100	50	100	94	100	92	80	83	91	100	100	69	91	85	97
19	94	93	80	92	50	90	76	87	67	70	67	73	100	93	77	78	77	85
21	19	47	40	58	20	30	29	47	25	40	17	55	73	33	23	22	32	40
28	0	0	0	8	0	0	6	7	0	20	0	0	0	0	8	0	3	4
31	31	60	20	42	20	90	47	73	50	70	42	64	73	80	85	44	39	67
Level of Performance*	Posttest only Subtest Performance:*																	
High (more than five items correct; score of 25 maximum)	21		25		33		20		11		18		22	20	9	5	14	
Low (fewer than three items correct; score of 11 or less)	29		0		0		10		33		27		11	7	36	47	20	

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the
in.

Unit II, Subtest 6, Food Chains

Several of the topic areas into which Unit II items were divided contain only one or two items. In such cases the label "subtest" has still been applied; however, the small number of items precludes much analysis as a subtest. This subtest contains two items. See Table 27 for relevant statistics, items, and results. Table 28 shows performance by class and also reports success of individual students in responding to both items.

One item (Item 6) dealt with the ability to identify the appropriate links in a food chain. Only 10 percent of the students were able to draw arrows to link the organisms appropriately in a simple food chain. The other item (Item 14) dealt with the recognition of the term "food web" as the appropriate term to describe the interrelationships

(Continued)

TABLE 27
Summary Information for Subtest 6
Food Chains and Webs
(two items)

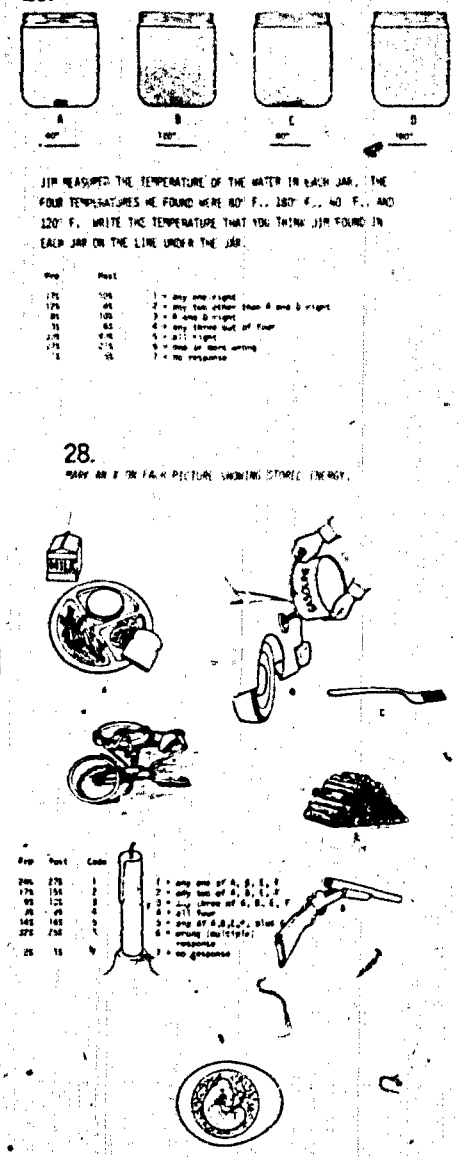
This pair of items assessed:
--the ability to identify the appropriate links in a food chain (Item 6)
--the recognition of the term "food web," for the interrelationships of a number of living things (Item 14)

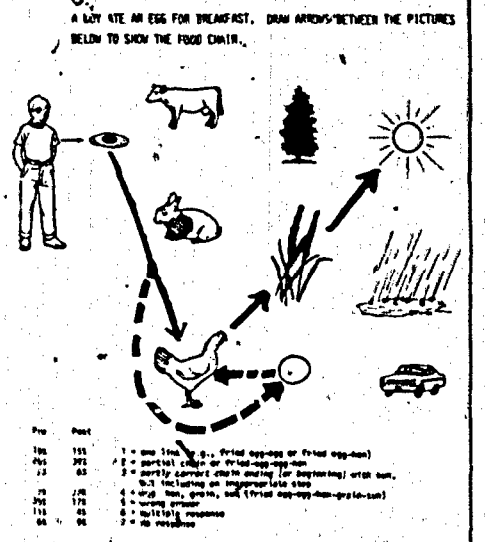
Subtest Weighted Scores:

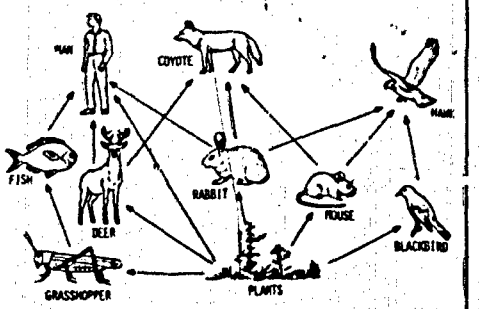
Item	Weights Per Option 1	2	3	4	5	6	7
6	0	0	3	5	0	0	0
14	0	0	5	0	-	0	0

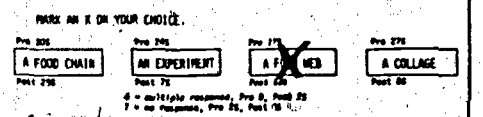
Subtest Statistics:

Maximum possible score: 10
Range of Scores: 0-10
Mean: 3
Standard Deviation: 3









of a number of living things. About half of the students correctly related the term to an illustration of a food web on the posttest. Considering both items as a subtest, only eight percent of the students answered more than one item correctly, while 42 percent answered neither item. In five of the ten classes, one third or more of the students were unsuccessful in answering either item (mean: 3, S.D.: 3, maximum score possible: 10).

A number of additions and revisions of the materials have been made in order to establish the concepts more effectively. In what is now Unit III, a new activity deals with what various animals eat. A food chain game has been revised to provide experiences in linking animals to things that they eat and ultimately to the sun. More experience related to food webs has been provided in several other activities. A new picture booklet deals with the interrelationships of plants and animals in an ecosystem, although "ecosystem" is not used in the materials. In Unit IV, eight activities have been developed relating to food chains and webs and the interrelationships of plants and animals in an ecosystem. This sequence of activities attempts to tie these relationships more closely to the students' own lives.

TABLE 28
Subtest 6 (Food Chains and Webs): Percent Choosing the Correct Response, by Class

Field Test Classes

Question N	00 Pre 16	00 Post 14	05 Pre 10	05 Post 4	06 Pre 10	06 Post 9	07 Pre 17	07 Post 10	08 Pre 12	08 Post 9	09 Pre 12	09 Post 11	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined Pre 77	All Classes Combined Post 111
6	6	0	0	0	0	30	0	13	0	0	8	0	9	0	8	4	1	10
14	13	73	0	58	20	90	29	73	25	60	17	9	18	33	54	35	17	53

Level of Performance — Posttest only Subtest Performance*

High (more than one item correct)	14	25	33	20	0	0	0	0	18	5	8
Low (none correct)	29	25	11	30	33	82	78	47	27	68	42

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the mean.

Unit II, Subtest 7, Food Energy

Five items were included in this subtest. Three of the items allowed partial credit for some responses. The two items which students found most difficult required a written response. See Table 29 for a summary of subtest statistics, items, and responses. Table 30 reports results by class.

Three of the items (Items 2, 8, and 24) assessed the knowledge that plants make their own food from sunlight and other materials. One of these three items (Item 2) required the students to state that plants make their own food or that they need air, water, soil, and sunlight to get their food. One fifth of the students stated this on the posttest, whereas only one percent could do so on the pretest. On the other two of these three items, about one half of the students marked a response that sunlight was the plant's source of energy, and that plants could change the sun's energy into food energy.

(Continued)

TABLE 29
Summary Information for Subtest 7
Food Energy
(five items)

This group of items measured:
--the knowledge that plants make their own food from sunlight and other materials (Items 2, 8, 24)
--the understanding that plants are the ultimate source of all food for all living things (Items 1 and 15)

Subtest Weighted Scores:

	Weights Per Option						
Item	1	2	3	4	5	6	7
1	1	1	1	3	5	0	0
2	1	1	3	5	5	0	0
8	0	5	0	0	-	0	0
15	0	5	0	0	-	0	0
24	0	0	0	5	-	0	0

Subtest Statistics:

Maximum possible score: 25
Range of Scores: 0-25
Mean: 11
Standard Deviation: 6

2.
WHERE DO PLANTS GET THEIR FOOD?

Pre	Post	
71%	19%	1. From air and/or water and/or soil.
2%	9%	2. From sunlight.
15%	40%	3. From all of the above.
3%	4%	4. They make their own food.
4	14%	5. Same as #4 but includes fact that they need sunlight, air, water, and soil to make their own food.
7%	3%	6. Wrong answer.
0%	0%	7. No response.

8.
KAREN PLANTED A BEAN SEED IN HER GARDEN. SHE WATERED THE PLANT EVERY DAY. AND AFTER SIX WEEKS THE BEAN WAS ONE FOOT TALL. WHERE DID THE PLANT GET THE ENERGY TO GROW THAT MUCH?

MARK AN X ON YOUR CHOICE.

1 SOIL	2 SUNLIGHT	3 AIR	4 WATER
Pre 20% Post 10%	Pre 41% Post 84%	Pre 6% Post 0%	Pre 29% Post 10%

6 = multiple response: Pre 5%, Post 0%
7 = no response: Pre 0%, Post 0%

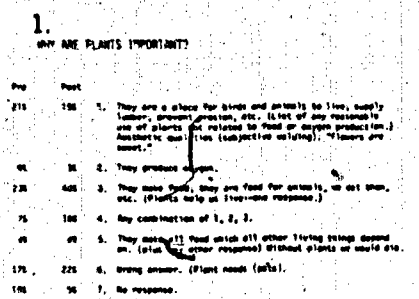
24.
WHICH LIVING THING CAN CHANGE THE SUN'S ENERGY DIRECTLY INTO FOOD ENERGY? MARK AN X ON YOUR CHOICE.

1 RABBIT	2 MUSHROOMS	3 COW	4 GRASS
Pre 0% Post 3%	Pre 14% Post 10%	Pre 39% Post 40%	Pre 47% Post 44%

6 = multiple response, Pre 1%, Post 2%
7 = no response, Pre 1%, Post 1%

Food Energy (Continued)

The remaining two items in this subtest (Items 1 and 15) assessed the understanding that plants are the ultimate source of all food for all living things, and that without plants we would die. Only six percent of the students expressed this answer in written form on Item 1; however, half of the students were able to identify plants as the things that the animals depended upon in Item 15, an illustration of a food web.



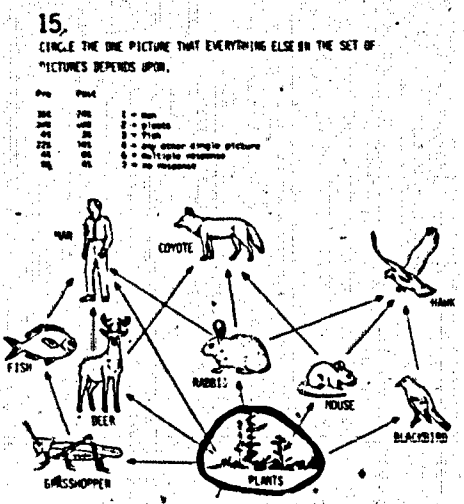


Table 30 reports the results by class for all five items combined. It shows the percentage of students whose performance was high (more than three items correct) or low (fewer than two items correct).

Considering the five items as a subtest, 10 percent of the students answered more than three of the items correctly, while 15 percent answered fewer than two items correctly (mean: 11, S.D.: 6, maximum score possible: 25). One third of the students in two of the classes were high performers on this subtest. Overall, the level of performance was reasonable considering the difficulty of the concepts.

A series of plant experiments were used to communicate many of the ideas in this subtest; these experiments have been considerably revised. The sequence of the materials has also been changed, in revision.

TABLE 30
Subtest 7 (Food Energy): Percent Choosing the Correct Response, by Class

	Field Test Classes																							
	00		05		06		07		08		09		11	12	13	14	All Classes Combined	
Question N	Pre 16	Post 14	Pre 10	Post 4	Pre 10	Post 9	Pre 17	Post 10	Pre 12	Post 9	Pre 12	Post 11	Post 9	Post 15	Post 11	Post 19	Pre 77	Post 111
1	0	13	10	0	0	0	0	0	0	10	0	0	19	7	0	4	4	6
2	0	40	0	0	0	60	0	0	0	0	0	0	36	0	46	0	1	18
8	0	73	30	33	30	80	35	33	33	50	33	45	73	87	82	35	40	56
15	19	53	20	58	40	60	41	53	17	10	33	45	64	47	77	26	26	48
24	50	40	40	50	50	90	41	40	8	10	33	36	91	47	59	30	37	44
Level of Performance	Posttest only Subtest Performance:*																	
High (more than three items correct)	36	0	44	0	11	0	33	7	27	0	10
Low (fewer than two items correct)	29	0	0	20	44	18	0	13	9	32	15

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the mean.

Unit II, Subtest 8, Weight and Temperature

Two items were included in this subtest. One of the items (Item 23) assessed an understanding of how a balance should be used. The other item (Item 30) served to educate students' general understanding of what numbers in the Fahrenheit scale mean. As in all subtests, no attempt was made to fully assess student understanding. Instead these two items merely sampled understandings of measurement and served as indicators rather than an inventory of what had been learned. Table 31 summarizes subtest statistics, and Table 32 shows student performance on both items combined, by class.

In only three classes did more than two thirds of the class succeed on Item 23, understanding the balance. On Item 30, room temperature, less than half the total group gave the correct response. In two classes, however, more than two thirds of the class were successful. Item 30 was drawn from science items used in the 1970 National Assessment of Educational Progress. Results for the nine- and 13-year old national sample are shown on the item. As can be noted, the response of the 13- to 14-year-old EMH field test group resembles the response of nine-year-old normal children on this item.

Considering the two items together as an index to student understanding of measurement, about one third of the total group of students answered both items correctly, while one fifth answered neither item correctly. For performance by class, one third or more of the students in five classes were high performers, while about one third of the students in two classes were unsuccessful. In only two classes were students predominantly successful on these items with no students unsuccessful.

In the revised materials, to save the time spent constructing balances, an inexpensive commercial balance has been selected; more time is devoted to learning to read the balance and to measuring different things.

TABLE 31
Summary Information for Subtest 8
Weight and Temperature
(two items)

These two items measured:
--understanding of a balance (Item 23)
--awareness of appropriate room temperature (Item 30)

Subtest Weighted Scores:

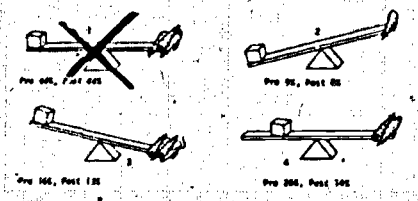
Item	Weights Per Option 1	2	3	4	5	6	7
23	5	0	0	0	-	0	0
30	0	0	5	0	-	0	0

Subtest Statistics:

Maximum possible score: 10
Range of Scores: 0-10
Mean: 5.4
Standard Deviation: 3.7

23.
JOHN CONSTRUCTED A BALANCE TO WEIGH FOOD FOR HIS PET RABBIT. ON ONE SIDE OF THE BALANCE HE PLACED A WEIGHT EQUAL TO THE AMOUNT OF FOOD HE WANTED FOR THE RABBIT. MARK AN X ON THE PICTURE WHICH SHOWS THE AMOUNT OF FOOD HE SHOULD GIVE THE RABBIT?

MARK AN X ON YOUR CHOICE.



30.
WHAT IS THE TEMPERATURE THAT IS THOUGHT MOST COMFORTABLE FOR A CLASSROOM?

MARK AN X ON YOUR CHOICE.

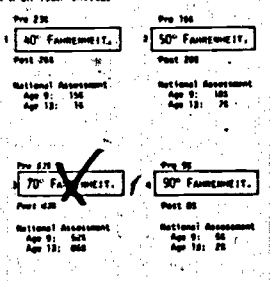


TABLE 32
Subtest 8 (Weight and Temperature): Percentage Choosing the Correct Response, by Class

Field Test Classes

Question N	00 Pre 16	00 Post 14	05 Pre 10	05 Post 4	06 Pre 10	06 Post 9	07 Pre 17	07 Post 10	08 Pre 12	08 Post 9	09 Pre 12	09 Post 11	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined Pre 77	All Classes Combined Post 111
23	44	53	10	58	80	100	35	87	83	50	67	73	46	60	31	57	48	64
30	44	33	40	25	30	50	8	80	50	70	42	36	36	20	0	35	51	42

Level of Performance

Posttest only Subtest Performance:*

High (both correct)	14	25	55	70	44	36	33	6	27	21	30
Low (neither correct)	29	25			22	27	44	20	18	31	21

*Percentage of students per class whose total subtest score was more than one Standard Deviation above or below the mean.

Unit II, Subtest 9, Graphing

While several items had been constructed to assess a student's understanding of graphing, defects in some of the items reduced the usable number of items to one. This item, drawn from science items used in the 1970 National Assessment of Educational Programs, assessed the ability to recognize the correct graphing of three items of information. Results are shown on the item for the total group in this field test and for the nine-year-old group sampled in the National Assessment study. Table 33 shows percentage of students by class choosing the best response to this item.

Sixty seven percent of the group correctly answered the item on the posttest, representing little change from the pretest. This result for 13- to 14-year-old children was clearly above the performance of nine-year-old normal children in the National Assessment sample. It is probably not much below what normal 13-year-old children can do. For four field test classes, about four out of five students answered the item correctly; half the students in three other classes were unsuccessful.

These results reflected a moderately high level of understanding. However, items in Subtests 1 and 3 indicated student difficulty with graph interpretation. Since graphs are used in various activities in the materials to summarize and compare data, it was considered necessary to develop more fully the concept of graphing. A new activity has been devised which provides practice in graphing and reading graphs of various kinds. Other activities which use graphs were expanded to provide more emphasis on the graphing process itself.

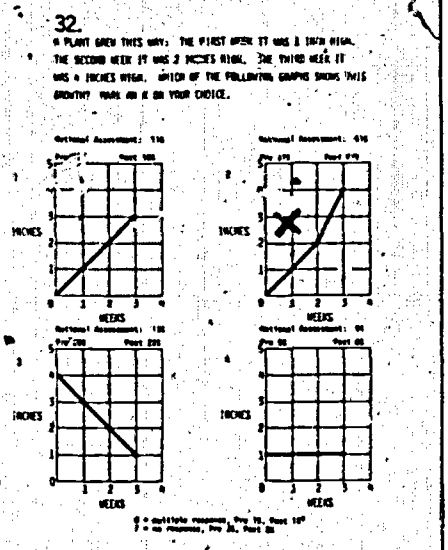


TABLE 33
Subtest 9 (Graphing): Percentage Choosing the Correct Response, by Class

Question N	Field Test Classes 00 Pre 14	00 Post 14	05 Pre 4	05 Post 4	06 Pre 9	06 Post 9	07 Pre 10	07 Post 10	08 Pre 9	08 Post 9	09 Pre 11	09 Post 11	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined Pre 111	All Classes Combined Post 111
32	69	64	30	50	40	90	71	80	67	63	42	82	78	53	45	68	57	67

Unit II, Subtest 10, Categorizing

This subtest consisted of two items which involved not only the process of grouping things but an understanding of specific concepts. One item (Item 7) assessed whether students recognized seeds as living things. The other (Item 35) required selection of a balanced meal from four pictured menus. Table 34 shows a summary of subtest statistics. Table 35 shows the percentage of students by class who chose the best response to each item.

TABLE 34
Summary Information for Subtest 10
Categorizing.
(two items)

This group of items assessed:
--recognition that seeds are living things (Item 7)
--recognition of a balanced meal (Item 35)

Subtest Weighted Scores:

	Weights Per Option						
Item	1	2	3	4	5	6	7
7	10	0	0	5	0	0	0
35	5	0	0	3	-	0	0

Subtest Statistics:

Maximum possible score: 10
Range of Scores: 0-10
Mean: 6.1
Standard Deviation: 3.2

These two items were answered correctly by approximately half of the field test group. No class stood out above the others in performance on these items. About one third of the students in the total group answered both items correctly. In interviews about Item 35, only one in four students (of a random sample) gave a reason for marking their answer the way they did and explained what a balanced meal was. The others either indicated the meal was one they liked to eat, that it gave them energy, or that there was more of it, and so on.

In the revised materials, a "Full and Healthy Game" was created to develop the concept of a balanced diet. An activity was also added which requires students to grow plants from seeds.

In summary, performance as measured by the 19 items related to Unit II of ME AND MY ENVIRONMENT was moderate. In six of the classes one third or more of the students were highly successful on at least one of the subtests. Considerable revision of Unit II materials should result in both a higher level of success and a greater number of students experiencing a justifiable degree of success.

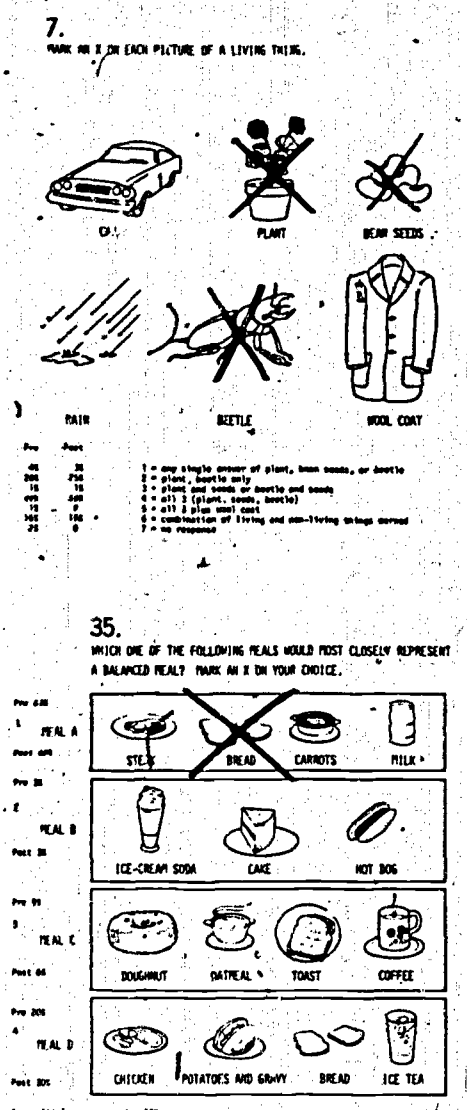


TABLE 35
Subtest 10 (Categorizing): Percentage Choosing the Correct Response, by Class

Question N	Field Test Classes 00 Pre 14	00 Post	05 Pre 4	05 Post	06 Pre 9	06 Post	07 Pre 10	07 Post	08 Pre 9	08 Post	09 Pre 11	09 Post	11 Post 9	12 Post 15	13 Post 11	14 Post 19	All Classes Combined Pre 111	All Classes Combined Post
7	56	33	50	50	50	60	41	47	50	40	25	45	73	67	46	74	49	56
35	38	40	30	25	60	80	71	60	50	50	58	45	46	33	54	57	51	49

BIOLOGICAL SCIENCES CURRICULUM STUDY

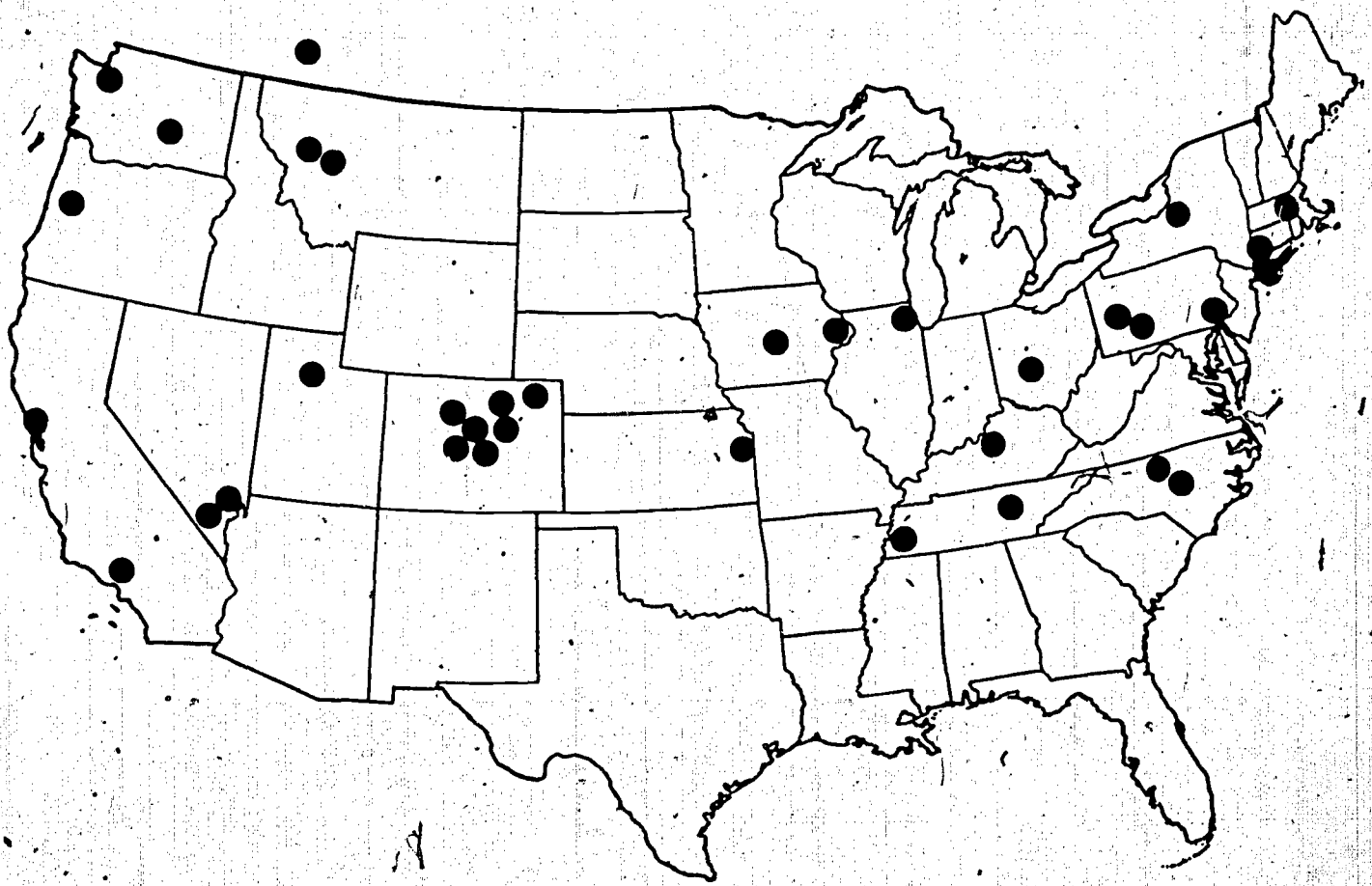
Me and my Environment

ACKNOWLEDGMENTS

No report on ME AND MY ENVIRONMENT would have been possible without the dedicated contributions of many teachers. Their help in writing and testing these materials has been invaluable in producing a successful and meaningful curriculum for EMH children.

The evaluation of materials has represented a joint effort of the staff consultants for the project. Special thanks are due to evaluation consultants Austin J. Conolly, Marlys Mitchell, and James Whorton who assisted the staff in generating hundreds of test items from which those reported herein were selected. Appreciation is also due to John B. Thornton for the design and formatting of this entire series of reports.

Me and My Environment



FIELD TEST TEACHERS

JUNE E. ALLBRANDT, *Lake Jr. High School, Denver, Colorado;* LARRY E. ALLEN, *Sentinel High School, Missoula, Montana;* VINCENT J. ALVINO, *Goleta Valley Jr. High School, Goleta, California;* ALICE B. BIGHAM, *Aycock Jr. High School, Raleigh, North Carolina;* DANNETTE BOYLE, *Nevin Platt Jr. High School, Boulder, Colorado;* CHAMINADE C. FARMER, *Gibson Jr. High School, Las Vegas, Nevada;* SHELDON J. FINE, *C. I. S. 148, Bronx, New York;* MARY L. FREEMAN, *Hodgkins Jr. High School, Westminster, Colorado;* CARL E. HOFF, *Lafayette Middle School, Lafayette, Colorado;* ALMA W. JENKINS, *Skiles Middle School, Evanston, Illinois;* JO ANN JEPPSON, *Granite Park Jr. High School, Salt Lake City, Utah;* STEVE G. JOHNSON, *Central School, Missoula, Montana;* WALLIS M. KIRYLUK, *Cole Jr. High School, Denver, Colorado;* MOLLIE P. KITE, *Pathfinder School, Bethel Park, Pennsylvania;* CANDACE J. LIGHT, *Hodgkins Jr. High School, Westminster, Colorado;* CECIL G. LINDER, *Roosevelt Jr. High School, Eugene, Oregon;* VITO LOMBARDO, *F. D. R. Middle School, Bristol, Pennsylvania;* RICHARD MATHIS, *Shadowlawn Learning Center, Arlington, Tennessee;* EDWARD H. McCANN, *Mon Valley School, West Mifflin, Pennsylvania;* MOLLY McCARTHY, *Benjamin Franklin Jr. High School, San Francisco, California;* R. DOUGLAS McCULLOUGH, *L. Y. Cairns Vocational School, Edmonton, Alberta, Canada;* EDWARD F. McNULTY, *Eastwood Jr. High School, Syracuse, New York;* JOAN A. McNULTY, *Central Jr. High School, DeWitt, Iowa;* MICHAEL W. RAE, *Granite Park Jr. High School, Salt Lake City, Utah;* EVA G. REEDY, *Benjamin Franklin Jr. High School, San Francisco, California;* TOM E. RODGERS, *Garside Jr. High School, Las Vegas, Nevada;* EDWARD H. SHERMAN, *I. S. 131, Bronx, New York;* HOWARD E. SHIPLEY, *Judson Hill School, Morristown, Tennessee;* EDITH D. SHIPMAN, *Pacific Prevocational School, Seattle, Washington;* BETTY S. SILVERTHORN, *Selah Jr. High School, Selah, Washington;* MARY D. SMITH, *Shepard Jr. High School, Durham, North Carolina;* JOSEPH SOUSA, *Harrington Way Jr. High School, Worcester, Massachusetts;* FRED W. STRICKLAND, *Nathan Weeks Jr. High School, Des Moines, Iowa;* HILDA S. THACH, *Pierson Jr. High School, Kansas City, Kansas;* SUSAN K. THOMASGARD, *Merrill Jr. High School, Denver, Colorado;* CARLA J. WATTS, *UNC Lab School, Greeley, Colorado;* MARY S. WHITE, *Shawnee Jr. High School, Louisville, Kentucky;* SUE WRIGHT, *West Locust Elementary School, Wilmington, Ohio;* AUGUST ZEITLOW, *UNC Lab School, Greeley, Colorado.*